

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 03-25-03
Searcher: Beverly E 4994
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
_____ STIC
_____ CM-1
_____ Pre-S
Type of Search
_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors
_____ IG Suite
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

1900-1901

1

1

1

1

1

1

1

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:53:49 ; Search time 12426 Seconds
(without alignments)
11562.897 Million cell updates/sec

Title: US-10-092-880-3

Perfect score: 4937
Sequence: 1 taaatacaagataataaa.....tttaacaggttattattatg 4937

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.ov.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4937	100.0	4937	6	AR032348	AR032348 Sequence
2	4937	100.0	4937	6	AR083955	AR083955 Sequence
3	4937	100.0	4937	6	AR145899	AR145899 Sequence
4	4937	100.0	4937	6	I25182	I25182 Sequence 3
5	4937	100.0	4937	6	I35771	I35771 Sequence 3
6	4935.4	100.0	4937	6	BD010356	BD010356 High molr
7	4900.2	99.3	8983	1	HIU08875	U08875 Haemophilus
8	4820.2	97.6	9323	6	AR032350	AR032350 Sequence
9	4820.2	97.6	9323	6	AR083957	AR083957 Sequence
10	4820.2	97.6	9323	6	AR145901	AR145901 Sequence
11	4820.2	97.6	9323	6	I25184	I25184 Sequence 6
12	4820.2	97.6	9323	6	I35773	I35773 Sequence 6
13	4817	97.6	9323	6	BD010358	BD010358 High molr
14	2689.8	54.5	9221	1	HIU08876	U08876 Haemophilus
15	2680.2	54.3	5116	6	AR032347	AR032347 Sequence
16	2680.2	54.3	5116	6	AR083954	AR083954 Sequence
17	2680.2	54.3	5116	6	AR145898	AR145898 Sequence
18	2680.2	54.3	5116	6	I25181	I25181 Sequence 1
19	2680.2	54.3	5116	6	I35770	I35770 Sequence 1
20	2567	52.0	9171	6	AR032349	AR032349 Sequence
21	2567	52.0	9171	6	AR083956	AR083956 Sequence
22	2567	52.0	9171	6	AR145900	AR145900 Sequence
23	2567	52.0	9171	6	I25183	I25183 Sequence 5
24	2567	52.0	9171	6	I35772	I35772 Sequence 5
25	2559.8	51.8	5056	6	BD010355	BD010355 High molr
26	2507	50.8	9111	6	BD010357	BD010357 High molr
27	2368.6	48.0	6976	1	AF180944	AF180944 Haemophil
28	2246.6	45.5	4803	6	AR083959	AR083959 Sequence
29	2245	45.5	4803	6	BD010360	BD010360 High molr
30	2202.2	44.6	4702	6	AR032352	AR032352 Sequence
31	2202.2	44.6	4702	6	AR145903	AR145903 Sequence
32	2202.2	44.6	4702	6	I25186	I25186 Sequence 8
33	2202.2	44.6	4702	6	I35775	I35775 Sequence 8
34	1974	40.0	4794	6	AR083958	AR083958 Sequence
35	1959.8	39.7	4795	6	BD010359	BD010359 High molr
36	1476.4	29.9	4287	6	AR032351	AR032351 Sequence
37	1476.4	29.9	4287	6	AR145902	AR145902 Sequence
38	1476.4	29.9	4287	6	I25185	I25185 Sequence 7
39	1476.4	29.9	4287	6	I35774	I35774 Sequence 7
C 40	94.6	1.9	10029	1	AE013292	AE013292 Methanosa
C 41	83.8	1.7	177899	10	AC123042	AC123042 Mus muscu
C 42	83.4	1.7	325808	2	AC121787	AC121787 Mus muscu
C 43	80.8	1.6	210821	10	AL731814	AL731814 Mouse DNA
C 44	78.4	1.6	146734	10	AL590415	AL590415 Mouse DNA
C 45	78.2	1.6	190380	2	AC123847	AC123847 Mus muscu

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1	AR032348	Sequence 3 from patent US 5869065.	AR032348	AR032348.1	GI:5947953	Unknown.	Unclassified.	1 (bases 1 to 4937)	Barenkamp, S.J. and St. Geme, J. William, III.	High molecular weight surface proteins of non-typeable haemophilus	Patent: US 5869065/A 3 09-FEB-1999;	Location/Qualifiers
AR032348	Sequence 3 from patent US 5869065.	AR032348	AR032348.1	GI:5947953	Unknown.	Unclassified.	1 (bases 1 to 4937)	Barenkamp, S.J. and St. Geme, J. William, III.	High molecular weight surface proteins of non-typeable haemophilus	Patent: US 5869065/A 3 09-FEB-1999;	Location/Qualifiers	

linear PAT 29-SEP-1999

Db 2041 GGTGTTTTTAATATTACCGCCGCTTCGTAGCTTTGAAGGTGGAATTAACAAAGCAGC 2100
QY 2101 GACGGGGCAATGCTAAATTTGTCGCCAGGCACTGTAAACCATTTACAGGAGAGGAAA 2160
Db 2101 GACGGGGCAATGCTAAATTTGTCGCCAGGCACTGTAAACCATTTACAGGAGAGGAAA 2160
QY 2161 GATTTACGGGCTAACACAGCTATCTTTAAACGGAAACGGGTGAAGGCTGAATATCATTTCA 2220
Db 2161 GATTTACGGGCTAACACAGCTATCTTTAAACGGAAACGGGTGAAGGCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTTAACCCACCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280
Db 2221 TCAGTGAATTAATTTAACCCACCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGCACCTGG 2340
Db 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGCACCTGG 2340
QY 2341 AAGCTCAGTGCTTAATCTAGNAGCAGCGGCAATTTTACCTTTATTAATATACATTTCA 2400
Db 2341 AAGCTCAGTGCTTAATCTAGNAGCAGCGGCAATTTTACCTTTATTAATATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGCTCTGCGGGGTGNAATTTTAACGGC 2460
Db 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGCTCTGCGGGGTGNAATTTTAACGGC 2460
QY 2461 GATAATGGCAACATGCTCAATCTCAAGAGAGGAGGAAAGTAAATTTCAATTAATAA 2520
Db 2461 GATAATGGCAACATGCTCAATCTCAAGAGAGGAGGAAAGTAAATTTCAATTAATAA 2520
QY 2521 CCAACGAGAACATGATCAACACAGCAACCTTTACCAATTCGGTTTATGACCAATATCA 2580
Db 2521 CCAACGAGAACATGATCAACACAGCAACCTTTACCAATTCGGTTTATGACCAATATCA 2580
QY 2581 GCCACTGGTGGGGCTCTGTGTTTTTGTATATATATGCCAACCATTTCTGGCAGAGGGCT 2640
Db 2581 GCCACTGGTGGGGCTCTGTGTTTTTGTATATATATGCCAACCATTTCTGGCAGAGGGCT 2640
QY 2641 GAGTTAAATAGTGAATTAATATCTCAACGGCGCTAAATTTTACCTTTAAATTCCTCAT 2700
Db 2641 GAGTTAAATAGTGAATTAATATCTCAACGGCGCTAAATTTTACCTTTAAATTCCTCAT 2700
QY 2701 GTTCGGCGCATGACGCTTTTAAATCAACAAAGACTTAAACCAATAATGCAACCAATCA 2760
Db 2701 GTTCGGCGCATGACGCTTTTAAATCAACAAAGACTTAAACCAATAATGCAACCAATCA 2760
QY 2761 AATTTACGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGCAAGCAATGCCATC 2820
Db 2761 AATTTACGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGCAAGCAATGCCATC 2820
QY 2821 AATTTCAACCTACACATATCCATTTCTGGGGGTAATGTACCCCTTGGTGACAAAACCTCA 2880
Db 2821 AATTTCAACCTACACATATCCATTTCTGGGGGTAATGTACCCCTTGGTGACAAAACCTCA 2880
QY 2881 AGCAGCAGCATTACGGGGAATATTTACCTATCGAGAAAGCAGCAATGTTACGCTAGAAGCC 2940
Db 2881 AGCAGCAGCATTACGGGGAATATTTACCTATCGAGAAAGCAGCAATGTTACGCTAGAAGCC 2940
QY 2941 AATACGCCCTTAATACGAAACATAAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000
Db 2941 AATACGCCCTTAATACGAAACATAAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTTAAGTTTAACCTGCGGAAATCCAGATTTAAAGCAATCTCACTATT 3060
Db 3001 GTTAATGGGAGTTTAAGTTTAACCTGCGGAAATCCAGATTTAAAGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACCGGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACCGGCAATTTT 3120
QY 3121 ACCAATAATGCACTGCGGCAATTAATATAACAAGAGAGTGTAAACCTTTGGCAATGTT 3180
Db 3121 ACCAATAATGCACTGCGGCAATTAATATAACAAGAGAGTGTAAACCTTTGGCAATGTT 3180

Db 3121 ACCAATAATGCACTGCGGCAATTAATATAACAAGAGAGTGTAAACCTTTGGCAATGTT 3180
QY 3181 ACCAATCATGTTGATTTTAAACATTTACCACATCACGCTAAACGCAACCAAGCAATCATC 3240
Db 3181 ACCAATCATGTTGATTTTAAACATTTACCACATCACGCTAAACGCAACCAAGCAATCATC 3240
QY 3241 GGCGGACATATAATCAACAAAAAGGAGCTTAAATATTACAGACAGTATAATATGATGCT 3300
Db 3241 GGCGGACATATAATCAACAAAAAGGAGCTTAAATATTACAGACAGTATAATATGATGCT 3300
QY 3301 GAAATCAAATTTGGCGCAATATCTCGCAAAAAGGCAACCTCACGATTTCTTCGGAT 3360
Db 3301 GAAATCAAATTTGGCGCAATATCTCGCAAAAAGGCAACCTCACGATTTCTTCGGAT 3360
QY 3361 AAAATTAATATCACCAACAGATTAACAAATCAAAAAGGATTTGATGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATATCACCAACAGATTAACAAATCAAAAAGGATTTGATGAGAGGACTCTAGT 3420
QY 3421 TCAGATGCGACAAAGTAAATGCGCAACCTTAACATTTAAAAACCAAGAAATTTGAAATTCACAGAA 3480
Db 3421 TCAGATGCGACAAAGTAAATGCGCAACCTTAACATTTAAAAACCAAGAAATTTGAAATTCACAGAA 3480
QY 3481 GACCTAAGTATTTTACAGGTTTCAATTAAGCAGAGATTTACAGCAAAAGATGTTAGAGATTTA 3540
Db 3481 GACCTAAGTATTTTACAGGTTTCAATTAAGCAGAGATTTACAGCAAAAGATGTTAGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTGCGGAAGCAAAAACAGTAATTTTAAAC 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTGCGGAAGCAAAAACAGTAATTTTAAAC 3600
QY 3601 AATGTTAAAGATTCAAAATCTCTGCTCAGCGTCACATGTCACATTAATAGCAAAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAATCTCTGCTCAGCGTCACATGTCACATTAATAGCAAAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGCGGCGAGTAAAGCAATACGCAACAGTACCGGCTTAATC 3720
Db 3661 AAAACATCTAGCAGCAATGCGGCGAGTAAAGCAATACGCAACAGTACCGGCTTAATC 3720
QY 3721 AATGTTCAAAAATTTAGAGTAAACAAAGATATTCTCTCTCAAAACAGTAATATC 3780
Db 3721 AATGTTCAAAAATTTAGAGTAAACAAAGATATTCTCTCTCAAAACAGTAATATC 3780
QY 3781 ACCGGCTCGGAAAGGTTTACCACACAGCGCTCGACCATTTAACGCAACAAATGGCAAA 3840
Db 3781 ACCGGCTCGGAAAGGTTTACCACACAGCGCTCGACCATTTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACAAACCAAAACAGTGTATATCAGCGGTACGATTTCCGGTAAACGCTAAGT 3900
Db 3841 GCAAGTATTACAAACCAAAACAGTGTATATCAGCGGTACGATTTCCGGTAAACGCTAAGT 3900
QY 3901 GTTAGCGGACGTGATTTAAACCAATTAACCACTAACTCCGCTCAAAATTTGAAGCAAAATCGGT 3960
Db 3901 GTTAGCGGACGTGATTTAAACCAATTAACCACTAACTCCGCTCAAAATTTGAAGCAAAATCGGT 3960
QY 3961 GAGGCTTAATGTAACAGTCCAAACAGTACAAATTTGGCGGTACAAATTTCCGGTAAATAGGTA 4020
Db 3961 GAGGCTTAATGTAACAGTCCAAACAGTACAAATTTGGCGGTACAAATTTCCGGTAAATAGGTA 4020
QY 4021 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATTTGGCGGAGAAATTAATCGGACA 4080
Db 4021 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATTTGGCGGAGAAATTAATCGGACA 4080
QY 4081 GAAGGAGCTGCAACCTTTAAACCCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTTAAACCCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140
QY 4141 ATCAGTTCAACTAAGGTCAGGTAGACCTCTTTGGCTCAGAAATGTTAGCATCGCAGGAAGC 4200
Db 4141 ATCAGTTCAACTAAGGTCAGGTAGACCTCTTTGGCTCAGAAATGTTAGCATCGCAGGAAGC 4200
QY 4201 ATTAAGTCTCTAATCTGACATTAATACTACAGGCACTTTAACCCGCTGCGAGGCTCG 4260
Db 4201 ATTAAGTCTCTAATCTGACATTAATACTACAGGCACTTTAACCCGCTGCGAGGCTCG 4260

QY	4261	GATATTAAAGCAACCGCGCACCTTGGTTATTAAAGCAAAAGATGCTAAGCTAAATGGT	4320
Db	4261	GATATTAAAGCAACCGCGCACCTTGGTTATTAAAGCAAAAGATGCTAAGCTAAATGGT	4320
QY	4321	GATGCATCAGGTGATAGTACAGAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG	4380
Db	4321	GATGCATCAGGTGATAGTACAGAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG	4380
QY	4381	ACTGGCGCAACCTCAAGCAGTGTGAATATCACTGCGGGATTTAAACACAGTAAATGGGTTA	4440
Db	4381	ACTGGCGCAACCTCAAGCAGTGTGAATATCACTGCGGGATTTAAACACAGTAAATGGGTTA	4440
QY	4441	AATATCATTTTGAAGATGGTGAAGAACACTGTGCGCTTAAAGAGGCAAGAAATTCAGGTG	4500
Db	4441	AATATCATTTTGAAGATGGTGAAGAACACTGTGCGCTTAAAGAGGCAAGAAATTCAGGTG	4500
QY	4501	AAATATATCCAGCCAGGTGTAGCAAGTGTAGAAGAGTAAATTAAGCGAAACGGTCTT	4560
Db	4501	AAATATATCCAGCCAGGTGTAGCAAGTGTAGAAGAGTAAATTAAGCGAAACGGTCTT	4560
QY	4561	GAAGAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAACTTTGGTGAAGT	4620
Db	4561	GAAGAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAACTTTGGTGAAGT	4620
QY	4621	GCTGTACGTTTTGTTGAGCCAAATTAATACAATTACAGTCAATACACAAATGAATTTACA	4680
Db	4621	GCTGTACGTTTTGTTGAGCCAAATTAATACAATTACAGTCAATACACAAATGAATTTACA	4680
QY	4681	ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTGAAGCGTGTCTCAAGTGGTAAAT	4740
Db	4681	ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTGAAGCGTGTCTCAAGTGGTAAAT	4740
QY	4741	GGCGCACGAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTAATTCACAAAG	4800
Db	4741	GGCGCACGAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTAATTCACAAAG	4800
QY	4801	GTAGATTCATCTTCGCAATGAAGTCATTTTATTTCTGATATTTTCTGATTTTCTGATTTT	4860
Db	4801	GTAGATTCATCTTCGCAATGAAGTCATTTTATTTCTGATATTTTCTGATTTTCTGATTTT	4860
QY	4861	GTTGAGTACGGGCTTTACCCATCTTTGAAAAAATTACGAGAAATACAAATTAAGTATTTT	4920
Db	4861	GTTGAGTACGGGCTTTACCCATCTTTGAAAAAATTACGAGAAATACAAATTAAGTATTTT	4920
QY	4921	AACAGGTTATTATTATG 4937	
Db	4921	AACAGGTTATTATTATG 4937	
RESULT 2			
LOCUS	AR083955	4937 bp	DNA linear PAT 01-SEP-2000
DEFINITION	Sequence 3 from patent US 5977336.		
ACCESSION	AR083955		
VERSION	AR083955.1 GI:10010726		
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 4937)		
AUTHORS	Barenkamp,S.J.		
TITLE	High molecular weight surface proteins of non-typeable haemophilus		
JOURNAL	Patent: US 5977336-A 3 02-NOV-1999;		
FEATURES	Location/Qualifiers		
source	1..4937		
BASE COUNT	1729 a	948 c	1009 g 1251 t
ORIGIN			
Query Watch	100.0%;	Score 4937;	DB 6; Length 4937;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 4937;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;

QY	1	TAAATATACAGATAATAAAAAATAAATCAAGATTTTTTGTGATGACAAACAACTTACAA	60
Db	1	TAAATATACAGATAATAAAAAATAAATCAAGATTTTTTGTGATGACAAACAACTTACAA	60
QY	61	CACCTTTTTTGCAGTCATATATGCAAAATATTTTAAAAAATAGTATAAATCCGCGCATATA	120
Db	61	CACCTTTTTTGCAGTCATATATGCAAAATATTTTAAAAAATAGTATAAATCCGCGCATATA	120
QY	121	AATGGTAAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT	180
Db	121	AATGGTAAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT	180
QY	181	CTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT	240
Db	181	CTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT	240
QY	241	CACATGAATGATGAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	300
Db	241	CACATGAATGATGAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	300
QY	301	GAACGCAAAATGATAAGTAATTTTAAATTTTCAACTAACCTTAGGAGAAATATGAACAAG	360
Db	301	GAACGCAAAATGATAAGTAATTTTAAATTTTCAACTAACCTTAGGAGAAATATGAACAAG	360
QY	361	ATATATCGTCTCAAAATTCAGCAAAACGCTGAATGCTTTGGTTCGCTGCTGCTGCTGCTG	420
Db	361	ATATATCGTCTCAAAATTCAGCAAAACGCTGAATGCTTTGGTTCGCTGCTGCTGCTGCTG	420
QY	421	CGGGTGTGTGACCATTCACAGAAAGGCTTCCGCTATGTTACTATCTTTTAGGTGTAAC	480
Db	421	CGGGTGTGTGACCATTCACAGAAAGGCTTCCGCTATGTTACTATCTTTTAGGTGTAAC	480
QY	481	CACCTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTTAGGTGTAAC	540
Db	481	CACCTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTTAGGTGTAAC	540
QY	541	CAATCTGTTTGTAGCAAGCGCTTACAAAGAAATGATGATGATGATGATGATGATGATGATG	600
Db	541	CAATCTGTTTGTAGCAAGCGCTTACAAAGAAATGATGATGATGATGATGATGATGATGATG	600
QY	601	CAAGTAGATGTTAAATAAACCAATTTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA	660
Db	601	CAAGTAGATGTTAAATAAACCAATTTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA	660
QY	661	CAATTTAATCATCGACCAAAATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720
Db	661	CAATTTAATCATCGACCAAAATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720
QY	721	GTATTTCAACCGTGTACATCTACCAAAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC	780
Db	721	GTATTTCAACCGTGTACATCTACCAAAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC	780
QY	781	GGACAAGTCTTTTAAATCAACCCAAATGATATCAAAATAGGTAAAGAGCGCAATTAATTAAC	840
Db	781	GGACAAGTCTTTTAAATCAACCCAAATGATATCAAAATAGGTAAAGAGCGCAATTAATTAAC	840
QY	841	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTACGAAACATCAAGCGCGGTAAAT	900
Db	841	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTACGAAACATCAAGCGCGGTAAAT	900
QY	901	TTACCTTTTCGAGCAACCAAAAGATAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	960
Db	901	TTACCTTTTCGAGCAACCAAAAGATAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	960
QY	961	ACTGTCGTTAAAGCGGAGTGAATCTTATGTTGGTGAAGTGAAGAGAGGAGGAGGAGGAG	1020
Db	961	ACTGTCGTTAAAGCGGAGTGAATCTTATGTTGGTGAAGTGAAGAGAGGAGGAGGAGGAG	1020
QY	1021	ATTAGCGTAAATGGTGGCAGCATTTCTTTTACTCGAGGCAAAAAAATCACCATCAGCGAT	1080
Db	1021	ATTAGCGTAAATGGTGGCAGCATTTCTTTTACTCGAGGCAAAAAAATCACCATCAGCGAT	1080

QY 1081 ATAATAAACCCCAACCACTTACTTACAGCATTCGCCGCGCTGAAATGAAGCGGTCAATCTG 1140
DB 1081 ATAATAAACCCCAACCACTTACTTACAGCATTCGCCGCGCTGAAATGAAGCGGTCAATCTG 1140
QY 1141 GCGCATATTTTGGCAAGCGGTAAACATTAATGTCGCTGCTGCACTATTCGAAACCAA 1200
DB 1141 GCGCATATTTTGGCAAGCGGTAAACATTAATGTCGCTGCTGCACTATTCGAAACCAA 1200
QY 1201 GGTAACTTTCTGCTGATTCTGTAAAGCAAGATATAAGCGGCAATATTTCTTCGCC 1260
DB 1201 GGTAACTTTCTGCTGATTCTGTAAAGCAAGATATAAGCGGCAATATTTCTTCGCC 1260
QY 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGG 1320
DB 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGG 1320
QY 1321 GCGAAGCTGATGATTACAGCGGATAAAGTACATTTAAAAACAGGTGCGAGTTATCGACCTT 1380
DB 1321 GCGAAGCTGATGATTACAGCGGATAAAGTACATTTAAAAACAGGTGCGAGTTATCGACCTT 1380
QY 1381 TCAGGTAAGAAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGGCTAAAGG 1440
DB 1381 TCAGGTAAGAAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGGCTAAAGG 1440
QY 1441 GGCATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGG 1500
DB 1441 GGCATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGG 1500
QY 1501 AAAGAAAAGCGGACCGGCTATTGTGTGGGGGATATTCGGTTAAATGACGGCAATATT 1560
DB 1501 AAAGAAAAGCGGACCGGCTATTGTGTGGGGGATATTCGGTTAAATGACGGCAATATT 1560
QY 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGTTTTGTGGAGACATCGGGCAT 1620
DB 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGTTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTATCCATGTACAGCAATGCAATTTGTTAAAAACAAAGAGTGGTGTAGACCCCTGAT 1680
DB 1621 TATTATCCATGTACAGCAATGCAATTTGTTAAAAACAAAGAGTGGTGTAGACCCCTGAT 1680
QY 1681 GATGTAAACAATTGAAGCGGAGAGCCCTTCGCAATATACCGGTATATAATGATGATTC 1740
DB 1681 GATGTAAACAATTGAAGCGGAGAGCCCTTCGCAATATACCGGTATATAATGATGATTC 1740
QY 1741 CCACAGCGCGGTGAGCAGGACCCCTAAAATAATAGCAACTCAAAACAGCCTA 1800
DB 1741 CCACAGCGCGGTGAGCAGGACCCCTAAAATAATAGCAACTCAAAACAGCCTA 1800
QY 1801 ACCAATACACTATTTCAAAATATCTGAAAACGCGTGGACAATGAATATAACGGCATCA 1860
DB 1801 ACCAATACACTATTTCAAAATATCTGAAAACGCGTGGACAATGAATATAACGGCATCA 1860
QY 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTTCTCCAT 1920
DB 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTTCTCCAT 1920
QY 1921 AGTAAAGGTACGCGTGGCGGAGGCTTCAGATTTGATGGAGATATTTACTTTAAAGGCGGA 1980
DB 1921 AGTAAAGGTACGCGTGGCGGAGGCTTCAGATTTGATGGAGATATTTACTTTAAAGGCGGA 1980
QY 1981 AATTTAAACCATTTATTTCTGGCGGATGGTTGATGTTCAATAAATATTTACGCTTTGATCAG 2040
DB 1981 AATTTAAACCATTTATTTCTGGCGGATGGTTGATGTTCAATAAATATTTACGCTTTGATCAG 2040
QY 2041 GGTATTTTAAATATTAACCGCGCTTCCGTAAGTCTTTGAGGTGGAAATATAACAGCAGC 2100
DB 2041 GGTATTTTAAATATTAACCGCGCTTCCGTAAGTCTTTGAGGTGGAAATATAACAGCAGC 2100
QY 2101 GACGCGCAATGCTAAATTTGTGCCCGGACCTGTAAACCTTACAGGAGAGGGA 2160
DB 2101 GACGCGCAATGCTAAATTTGTGCCCGGACCTGTAAACCTTACAGGAGAGGGA 2160
QY 2161 GATTTTCAGGGCTAACACGCTATCTTTAAACGGAACGGGTAAAGGCTGTAATATCATTTTCA 2220

DB 2161 GATTTTCAGGGCTAACACGCTATCTTTAAACGGAACGGGTAAAGGCTGAATATCATTTTCA 2220
QY 2221 TCAGTCAATAATTAACCCACATCTTACTGGCACAAATTAACATATCTCTGGGAATAAACA 2280
DB 2221 TCAGTCAATAATTAACCCACATCTTACTGGCACAAATTAACATATCTCTGGGAATAAACA 2280
QY 2281 ATTAACAACACTACGAGAAGAACACCTCGTATTCGCAACAGCAGCATATTCGCACTGG 2340
DB 2281 ATTAACAACACTACGAGAAGAACACCTCGTATTCGCAACAGCAGCATATTCGCACTGG 2340
QY 2341 AAGCTCAGTGTCTTAACTTAGAGACAGCGGCAAAATTTTACCTTTATTAATACATTTCA 2400
DB 2341 AAGCTCAGTGTCTTAACTTAGAGACAGCGGCAAAATTTTACCTTTATTAATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAACTCTCGAGGGGTGAATTTTAAACGCG 2460
DB 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAACTCTCGAGGGGTGAATTTTAAACGCG 2460
QY 2461 GTAAATGGCAACATGTCTTCAATCTCAAGAGAGGAGGAAAGTTAATTTCAAAATFAAA 2520
DB 2461 GTAAATGGCAACATGTCTTCAATCTCAAGAGAGGAGGAAAGTTAATTTCAAAATFAAA 2520
QY 2521 CCACAGCAGCAACATGACACCAACACCTTTACCAATTTCCGTTTTTACCAATATCACA 2580
DB 2521 CCACAGCAGCAACATGACACCAACACCTTTACCAATTTCCGTTTTTACCAATATCACA 2580
QY 2581 GCCACTGTGGGGGCTGTCTTTTTTTTGTATATATATATATATATATATATATATATATAT 2640
DB 2581 GCCACTGTGGGGGCTGTCTTTTTTTTGTATATATATATATATATATATATATATATAT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGGCTAAATTTTACCTTTAAATTTCCCAT 2700
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGGCTAAATTTTACCTTTAAATTTCCCAT 2700
QY 2701 GTTCGGGGGCTGATGACGCTTTTAAAAATCAACAAAGCTTTAACCAATAAATGCAACCAATTC 2760
DB 2701 GTTCGGGGGCTGATGACGCTTTTAAAAATCAACAAAGCTTTAACCAATAAATGCAACCAATTC 2760
QY 2761 AATTTTCAGCCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCGATC 2820
DB 2761 AATTTTCAGCCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCGATC 2820
QY 2821 AATTTACACTACACATATCTCTGGGGTAAATGCTACCTTGTGGACAAACTCA 2880
DB 2821 AATTTACACTACACATATCTCTGGGGTAAATGCTACCTTGTGGACAAACTCA 2880
QY 2881 AGCAGACGATTTACGGGAATATTTACTATCGAGAAAGCAGCAATGTTACGCTAGAGCC 2940
DB 2881 AGCAGACGATTTACGGGAATATTTACTATCGAGAAAGCAGCAATGTTACGCTAGAGCC 2940
QY 2941 AATAAGCGCCCTAATCAGCAAAAACATAAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000
DB 2941 AATAAGCGCCCTAATCAGCAAAAACATAAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTAAGTTTAACTGGCAAAATCGAGATATTAAGGCAATCTCAGCTATT 3060
DB 3001 GTTAATGGGAGTTAAGTTTAACTGGCAAAATCGAGATATTAAGGCAATCTCAGCTATT 3060
QY 3061 TCAGAAAGCGCCCTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120
DB 3061 TCAGAAAGCGCCCTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120
QY 3121 ACCAATAATGCGACTCGCGAAATTAATATAACACAGGAGTGTAAAACCTTTGGCAATGTT 3180
DB 3121 ACCAATAATGCGACTCGCGAAATTAATATAACACAGGAGTGTAAAACCTTTGGCAATGTT 3180
QY 3181 ACCAATGATGGTGAATTAACATTTACCCTACGCTTAACCGCAACCAAGAGCATCATC 3240
DB 3181 ACCAATGATGGTGAATTAACATTTACCCTACGCTTAACCGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAAAGGAGCTTAAATATTTACAGACAGTATAATATGATGCT 3300
DB 3241 GCGGAGATATAATCAACAAAAAGGAGCTTAAATATTTACAGACAGTATAATATGATGCT 3300

```
Db 3241 GCGGAGATATATCAACAAAAAGGAAGCTTAAATATTACAGACAGTAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGAAAAAGAGGCAACCTCACGATTTCTTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGCAATATCTCGAAAAAGAGGCAACCTCACGATTTCTTCCGAT 3360
QY 3361 AAAATTAATATCACCACACAGATACAAATCAAAAGGGTATTGATGGAGAGACTCTAGT 3420
Db 3361 AAAATTAATATCACCACACAGATACAAATCAAAAGGGTATTGATGGAGAGACTCTAGT 3420
QY 3421 TCAGATGGACAGTAAGTAAAGCAACCTCACTATTAAACCAAAAGAAATGAAATTTGACAGAA 3480
Db 3421 TCAGATGGACAGTAAGTAAAGCAACCTCACTATTAAACCAAAAGAAATGAAATTTGACAGAA 3480
QY 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
Db 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGTGCAGGCGGAGCCAAACAGTAATTTTAC 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGTGCAGGCGGAGCCAAACAGTAATTTTAC 3600
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCACAATGTGACACTAAATAGCAAAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCACAATGTGACACTAAATAGCAAAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAAGCAATAGCGCAACAGTACCGGCTTAAT 3720
Db 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAAGCAATAGCGCAACAGTACCGGCTTAAT 3720
QY 3721 ATTACTGCAAAAATGTAGAACTAAACAAAGATATTACTTCTCAAAACAGTAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAACTAAACAAAGATATTACTTCTCAAAACAGTAATATC 3780
QY 3781 ACCGGTCGGAAGGTTTACCACACAGCGCTCGACCATTAACGCAACAAATGGCAAA 3840
Db 3781 ACCGGTCGGAAGGTTTACCACACAGCGCTCGACCATTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACACCAACAGGTTGATATCAGCGGTAGCATTTCCGGTAAACACGGTAAGT 3900
Db 3841 GCAAGTATTACACCAACAGGTTGATATCAGCGGTAGCATTTCCGGTAAACACGGTAAGT 3900
QY 3901 GTTAGCGGACTGTGATTAAACCACTAAATCCGGCTCAAAATTTGAAGCAAAATCGGCT 3960
Db 3901 GTTAGCGGACTGTGATTAAACCACTAAATCCGGCTCAAAATTTGAAGCAAAATCGGCT 3960
QY 3961 GAGGCTAATGTAACAAAGTCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGGCTAATGTAACAAAGTCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020
QY 4021 AATGTTACGGCAACAGCTGGCGATTACAGTTGGGAATGGCGGAGAAATTAATGCGACA 4080
Db 4021 AATGTTACGGCAACAGCTGGCGATTACAGTTGGGAATGGCGGAGAAATTAATGCGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGCTCAGAATGGTAGCATCGCAGGAAGC 4200
Db 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGCTCAGAATGGTAGCATCGCAGGAAGC 4200
QY 4201 ATTAATGCTGCTAATGTGACATTAATTAATCAAGCAGCTTAACACCGTGGCAGGCTCG 4260
Db 4201 ATTAATGCTGCTAATGTGACATTAATTAATCAAGCAGCTTAACACCGTGGCAGGCTCG 4260
QY 4261 CATATTAAAGCAACAGCGGACCTTGTTTATTAAACGAAAGATGCTTAAGCTAAATGGT 4320
Db 4261 CATATTAAAGCAACAGCGGACCTTGTTTATTAAACGAAAGATGCTTAAGCTAAATGGT 4320
QY 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCACTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCACTCAACGCAAGCGGCTCTGGTAGTGTG 4380
```

```
QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATACCTGCGGATTAAACACAGTAAATGGGTTA 4440
Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATACCTGCGGATTAAACACAGTAAATGGGTTA 4440
QY 4441 AATATCATTTCCGAAAGATGGTGAACACACTGTGCGCTTAAAGAGCAAGGAAATTTAGGTG 4500
Db 4441 AATATCATTTCCGAAAGATGGTGAACACACTGTGCGCTTAAAGAGCAAGGAAATTTAGGTG 4500
QY 4501 AAATATATCCAGCCAGGTAGCAAGTGTAGAGAAGTAAATGAACGGAACCGCTCCTT 4560
Db 4501 AAATATATCCAGCCAGGTAGCAAGTGTAGAGAAGTAAATGAACGGAACCGCTCCTT 4560
QY 4561 GAAAAGTAAAAAGATTATCTGATGAAGAAAGAGAAACATTTAGCTAAACTTGGTGTAAAGT 4620
Db 4561 GAAAAGTAAAAAGATTATCTGATGAAGAAAGAGAAACATTTAGCTAAACTTGGTGTAAAGT 4620
QY 4621 GCTGTACGTTTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAAATGAATTTACA 4680
Db 4621 GCTGTACGTTTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAAATGAATTTACA 4680
QY 4681 ACCAGACCGTCAAGTCAAGTGATAATTTCTGAAGGTAAGCGGTGTTTCTCAAGTGGTAAT 4740
Db 4681 ACCAGACCGTCAAGTCAAGTGATAATTTCTGAAGGTAAGCGGTGTTTCTCAAGTGGTAAT 4740
QY 4741 GCGCACGAGTATGTACCAATGTTGCTGACGATGGACACCGGTAGTCAATTTGACAAG 4800
Db 4741 GCGCACGAGTATGTACCAATGTTGCTGACGATGGACACCGGTAGTCAATTTGACAAG 4800
QY 4801 GTAGATTTCATCCTCAATGAAGTCAATTTATTTTCGTTATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTCATCCTCAATGAAGTCAATTTATTTTCGTTATTTACTGTGGGTTAAA 4860
QY 4861 GTTCAAGTACGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAAATAAGTATTTT 4920
Db 4861 GTTCAAGTACGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAAATAAGTATTTT 4920
QY 4921 AACAGGTTATTATTATG 4937
Db 4921 AACAGGTTATTATTATG 4937
```

```
RESULT 3
ARI45899
LOCUS ARI45899
DEFINITION Sequence 3 from patent US 6218141.
ACCESSION ARI45899
VERSION ARI45899.1 GI:15109088
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4937)
AUTHORS Barenkamp,S.J.
TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 6218141-A 3 17-APR-2001;
FEATURES Location/Qualifiers
source 1. 4937
BASE COUNT 1729 a 948 c 1009 g 1251 t
ORIGIN
```

```
Query Match 100.0%; Score 4937; DB 6; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TAAATATACAGATAATAAATAAATCAAGATTTTGTGTGACAAACAAACAATTTACAA 60
Db 1 TAAATATACAGATAATAAATAAATCAAGATTTTGTGTGACAAACAAACAATTTACAA 60
QY 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
Db 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
```

Qy 121 AATGGTAAATCTTTTCATCTTTTCATCTTTAACTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
Db 121 AATGGTAAATCTTTTCATCTTTTCATCTTTAACTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
Qy 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Db 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Qy 241 CACATGAAATGATGAACCGAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 241 CACATGAAATGATGAACCGAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Qy 301 GAACGCAAAATGATAAAGTAAATTTAAATTTGTTCAACTAACCTTAGGAGAGAAATATGAACAAG 360
Db 301 GAACGCAAAATGATAAAGTAAATTTAAATTTGTTCAACTAACCTTAGGAGAGAAATATGAACAAG 360
Qy 361 ATATATCGTCTCAAAATTCAGCAAAACCGCTGAATGCTTTGGTCTGCTGCTGAATTTGGCA 420
Db 361 ATATATCGTCTCAAAATTCAGCAAAACCGCTGAATGCTTTGGTCTGCTGCTGAATTTGGCA 420
Qy 421 CGGGGTTGTGACCAATTCACAGAAAAGGCTTCGCGTATGTTACTATCTTTTAGTGTAAC 480
Db 421 CGGGGTTGTGACCAATTCACAGAAAAGGCTTCGCGTATGTTACTATCTTTTAGTGTAAC 480
Qy 481 CACTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTTAGTGTAACATCTTATTTCCA 540
Db 481 CACTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTTAGTGTAACATCTTATTTCCA 540
Qy 541 CAATCTGTTTAAAGGAGGCTTTACAGGAATGGATGTAGTACACGGCACAGCCACCTATG 600
Db 541 CAATCTGTTTAAAGGAGGCTTTACAGGAATGGATGTAGTACACGGCACAGCCACCTATG 600
Qy 601 CAAGTAGATGTTAAATTAACCAATATCCGCAACAGTGTGACGCTATCATTAATTTGGA 660
Db 601 CAAGTAGATGTTAAATTAACCAATATCCGCAACAGTGTGACGCTATCATTAATTTGGA 660
Qy 661 CAATTTACATCGACCAAAATGAATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
Db 661 CAATTTACATCGACCAAAATGAATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
Qy 721 GTATTTACCGTGTGTACATCTAACCAATCTCCCAATTTAAAGGATTTTAGATTTCTAAC 780
Db 721 GTATTTACCGTGTGTACATCTAACCAATCTCCCAATTTAAAGGATTTTAGATTTCTAAC 780
Qy 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATAGTAAAGAGCGCAATTTATTAAC 840
Db 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATAGTAAAGAGCGCAATTTATTAAC 840
Qy 841 ACTAATGCGTTTACGCGTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900
Db 841 ACTAATGCGTTTACGCGTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900
Qy 901 TTTACCTTCGAGCAAAACCAAGATTAAGCGCTGCTGAAATTTGTAATCAACGGTTTAAT 960
Db 901 TTTACCTTCGAGCAAAACCAAGATTAAGCGCTGCTGAAATTTGTAATCAACGGTTTAAT 960
Qy 961 ACTGTCGTTAAAGCGGAGTGAATCTTTTACGCAAAAGTGAAGAGGAGGAGGAGGAGGAGGAGG 1020
Db 961 ACTGTCGTTAAAGCGGAGTGAATCTTTTACGCAAAAGTGAAGAGGAGGAGGAGGAGGAGGAGG 1020
Qy 1021 ATTACGTTAAATGTTGGAGCAATTTCTTTTACGCGAGGCAAAATCACCATCAGCGAT 1080
Db 1021 ATTACGTTAAATGTTGGAGCAATTTCTTTTACGCGAGGCAAAATCACCATCAGCGAT 1080
Qy 1081 ATAATAAACCCCAACCAATTTACAGCATTTGCCGCGCTGAAATGAAGCGGCTCAATCTG 1140
Db 1081 ATAATAAACCCCAACCAATTTACAGCATTTGCCGCGCTGAAATGAAGCGGCTCAATCTG 1140
Qy 1141 GCGGATATTTTGGCAAAAGGCGGTAAACATTAATGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 1141 GCGGATATTTTGGCAAAAGGCGGTAAACATTAATGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200

Qy 1201 GGTAAACTTTCTGCTGATTTCTGAAGCAAGATAAAGCGGCAATATTTCTTCTTCCGCC 1260
Db 1201 GGTAAACTTTCTGCTGATTTCTGAAGCAAGATAAAGCGGCAATATTTCTTCTTCCGCC 1260
Qy 1261 AAAGAGGTGAAGCGGAAATTTGGCGTGAATTTCCGCTCAAAATCAGCAAGCTTAAAGC 1320
Db 1261 AAAGAGGTGAAGCGGAAATTTGGCGTGAATTTCCGCTCAAAATCAGCAAGCTTAAAGC 1320
Qy 1321 GGCAGCTGATGATTATACAGCGGATAAAGTCACATTTAAAAAACAGGTGAGTTATCGACCTT 1380
Db 1321 GGCAGCTGATGATTATACAGCGGATAAAGTCACATTTAAAAAACAGGTGAGTTATCGACCTT 1380
Qy 1381 TCAGGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
Db 1381 TCAGGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
Qy 1441 GGCATTTCAATTTAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db 1441 GGCATTTCAATTTAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Qy 1501 AAAGAGG 1560
Db 1501 AAAGAGG 1560
Qy 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAACCCGGTGGTGGTGGGAGACATCGGGGAT 1620
Db 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAACCCGGTGGTGGTGGGAGACATCGGGGAT 1620
Qy 1621 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTGGTGGTGGTGGTGGTGG 1680
Db 1621 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTGGTGGTGGTGGTGGTGG 1680
Qy 1681 GATGTAACTTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Db 1681 GATGTAACTTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Qy 1741 CCAACAGG 1800
Db 1741 CCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
Qy 1801 ACCAATCACTATTTCAATTTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
Db 1801 ACCAATCACTATTTCAATTTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
Qy 1861 AGAAACTTACCGTTAATAGCTCAATCACTCAATCACTCAATCACTCAATCACTCAATCACTCAAT 1920
Db 1861 AGAAACTTACCGTTAATAGCTCAATCACTCAATCACTCAATCACTCAATCACTCAATCACTCAAT 1920
Qy 1921 AGTAAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
Db 1921 AGTAAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Qy 1981 AATTTAACTTTATTTCTGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2040
Db 1981 AATTTAACTTTATTTCTGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2040
Qy 2041 GGTTTTTTAAATTTATTTCTGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2100
Db 2041 GGTTTTTTAAATTTATTTCTGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2100
Qy 2101 GAGCGCAATCTAAATTTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
Db 2101 GAGCGCAATCTAAATTTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
Qy 2161 GATTTAGG 2220
Db 2161 GATTTAGG 2220
Qy 2221 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Db 2221 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Qy 2281 ATTAACCAAACTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340

|||||
Db 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACACGCCATGATGCGCACTGG 2340
QY 2341 AAGCTCAGTCTCTTAATCTAGACAGCGCAAAATTTTACCCTTTATTAATAATACATTCA 2400
|||||
Db 2341 AAGCTCAGTCTCTTAATCTAGACAGCGCAAAATTTTACCCTTTATTAATAATACATTCA 2400
QY 2401 AGCAATAGCAAAAGCTTTAACAACACAGTATAGAAGCTCTCGAGGGTGAATTTTAAAGGC 2460
Db 2401 AGCAATAGCAAAAGCTTTAACAACACAGTATAGAAGCTCTCGAGGGTGAATTTTAAAGGC 2460
QY 2461 GTAATGGCAACATGTCATTCAATCTCAAGAAGAGCGCAAGTAAATTTCAAAATTAATA 2520
Db 2461 GTAATGGCAACATGTCATTCAATCTCAAGAAGAGCGCAAGTAAATTTCAAAATTAATA 2520
QY 2521 CCAACAGGAACATGAACCAACCAACCTTTTACCAAATTCGGTTTTAGCCAATATCACA 2580
Db 2521 CCAACAGGAACATGAACCAACCAACCTTTTACCAAATTCGGTTTTAGCCAATATCACA 2580
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCGCAACCATTTCTGGCAGAGGGCT 2640
Db 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCGCAACCATTTCTGGCAGAGGGCT 2640
QY 2641 GAGTTAAAATGAGTCAAAATTAATATCTCTAAGCGCGCTAAATTTTACCCTTAATTTCCCAT 2700
Db 2641 GAGTTAAAATGAGTCAAAATTAATATCTCTAAGCGCGCTAAATTTTACCCTTAATTTCCCAT 2700
QY 2701 GTTCGGGGGATGACGCTTTTAAAATCAACAAGAGCTTAAACCATAAATGCAACCAATTTCA 2760
Db 2701 GTTCGGGGGATGACGCTTTTAAAATCAACAAGAGCTTAAACCATAAATGCAACCAATTTCA 2760
QY 2761 AATTTTCAGCCTCAGACAGACGAAAGATGATTTTATGACGGGTACGCGCAATGGCATC 2820
Db 2761 AATTTTCAGCCTCAGACAGACGAAAGATGATTTTATGACGGGTACGCGCAATGGCATC 2820
QY 2821 AATTTCAACCTACAACATATCCATTTCTGGCGGTAAATGTTCACCTTGGTGACAAACTCA 2880
Db 2821 AATTTCAACCTACAACATATCCATTTCTGGCGGTAAATGTTCACCTTGGTGACAAACTCA 2880
QY 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940
Db 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940
QY 2941 AATAACGCCCTTAATCAGCAAAACATAGGGATAGAGTATATAAACTTGGCAGCTTGCTC 3000
Db 2941 AATAACGCCCTTAATCAGCAAAACATAGGGATAGAGTATATAAACTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATGCAGATATTAAAGGCAATCTCACTATT 3060
Db 3001 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATGCAGATATTAAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGGCCACTTTTAAAGGAAGAGTACGATACCTTAATATCACCGGCAATTTT 3120
Db 3061 TCAGAAAGGCCACTTTTAAAGGAAGAGTACGATACCTTAATATCACCGGCAATTTT 3120
QY 3121 ACCAATAATGGCAGTCCGCAATTAATATACACAGGAGTGGTAAACTTTGGCAATGTT 3180
Db 3121 ACCAATAATGGCAGTCCGCAATTAATATACACAGGAGTGGTAAACTTTGGCAATGTT 3180
QY 3181 ACCAATGATGGTATTAACATTTACCCTCACGCTAAACGCAACCAAGAGCATCATC 3240
Db 3181 ACCAATGATGGTATTAACATTTACCCTCACGCTAAACGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAAGAGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
Db 3241 GCGGAGATATAATCAACAAAAGAGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGCAACCTCACGATTTCTTCGGAT 3360
Db 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGCAACCTCACGATTTCTTCGGAT 3360
QY 3361 AAAATTAATATCAACCAACAGATACAAATCAAAAGGGTATTGATGGAGAGGACTCTAGT 3420
|||||

Db 3361 AAAATTAATATCAACCAACAGATAAACAATCAAAAAGGTTATTGATGGAGAGGACTCTAGT 3420
QY 3421 TCAGATCGGCAAGTAATGCCAACCTTAACCTATTATAAACCAAGAAATTTGAAATTCACAGAA 3480
Db 3421 TCAGATCGGCAAGTAATGCCAACCTTAACCTATTATAAACCAAGAAATTTGAAATTCACAGAA 3480
QY 3481 GACCTAAGTATTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
Db 3481 GACCTAAGTATTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAAATGACGGTAAACAGCGTGCAGCAAGCAAAACAGTAACTTTTAAAC 3600
Db 3541 ACTATTGGCAACAGTAAATGACGGTAAACAGCGTGCAGCAAGCAAAACAGTAACTTTTAAAC 3600
QY 3601 AATGTTTAAAGATTCAAAAATCTCTGCTGACGGTCAACAATGTGACACTTAAATAGCAAAAGTG 3660
Db 3601 AATGTTTAAAGATTCAAAAATCTCTGCTGACGGTCAACAATGTGACACTTAAATAGCAAAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGCGGACGTGAAGCAATAGCGACACACGATACCGGCTTAACT 3720
Db 3661 AAAACATCTAGCAGCAATGCGGACGTGAAGCAATAGCGACACACGATACCGGCTTAACT 3720
QY 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAAATATC 3780
QY 3781 ACCGCGTCGGAAGGTTACCACACAGCAGCTCGACCATTAACGCAACAAATGGCAAA 3840
Db 3781 ACCGCGTCGGAAGGTTACCACACAGCAGCTCGACCATTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACACGGTAAAGT 3900
Db 3841 GCAAGTATTACACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACACGGTAAAGT 3900
QY 3901 GTTAGCGCACTGGTGATTTTACCCTTAATCCGGCTCAAAAATTAAGCGAAATCGGGT 3960
Db 3901 GTTAGCGCACTGGTGATTTTACCCTTAATCCGGCTCAAAAATTAAGCGAAATCGGGT 3960
QY 3961 GAGGCTTAATTAACAAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATCGGTA 4020
Db 3961 GAGGCTTAATTAACAAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATCGGTA 4020
QY 4021 AATGTTACGGCAACCGCTGGCGATTAAACAGTTGGGAATGGCGCAGAAAATTAATGGGACA 4080
Db 4021 AATGTTACGGCAACCGCTGGCGATTAAACAGTTGGGAATGGCGCAGAAAATTAATGGGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGTAATCTGACTACTACGACCGGCTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGTAATCTGACTACTACGACCGGCTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGGTCAGGTAGACCTTTGGCTCAGAATGGTACGATCGCAGGAAGC 4200
Db 4141 ATCACTTCAACTAAGGGTCAGGTAGACCTTTGGCTCAGAATGGTACGATCGCAGGAAGC 4200
QY 4201 ATTATGCTGCTAATGTGACATTAATTAACAGGACCTTAACCACTTGGCAGGCTCG 4260
Db 4201 ATTATGCTGCTAATGTGACATTAATTAACAGGACCTTAACCACTTGGCAGGCTCG 4260
QY 4261 GATATTAAGCAACCAAGCGGCACTTTGTTTATTACGCAAAAAGATGCTAAGCTAAATGCT 4320
Db 4261 GATATTAAGCAACCAAGCGGCACTTTGTTTATTACGCAAAAAGATGCTAAGCTAAATGCT 4320
QY 4321 GATGATCAGGTGATGATACAGAACTGAATGATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4380
Db 4321 GATGATCAGGTGATGATACAGAACTGAATGATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4380
QY 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440
Db 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440
QY 4441 AATATCATTTTGAAGATGGTAGAAACACTGTGCCCTTAAAGAGGCAAGCAATTCAGGTG 4500
Db 4441 AATATCATTTTGAAGATGGTAGAAACACTGTGCCCTTAAAGAGGCAAGCAATTCAGGTG 4500

QY 1321 GCGAAGCTGATGATTACAGCGGATAAGTACATTAATAACAGTGCAGTTATCGACCTT 1380
DB 1321 GCGAAGCTGATGATTACAGCGGATAAGTACATTAATAACAGTGCAGTTATCGACCTT 1380
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACAGCGCGGAGGTAATAAAC 1440
DB 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACAGCGCGGAGGTAATAAAC 1440
QY 1441 GGCATTCAATTAGCAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500
DB 1441 GGCATTCAATTAGCAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500
QY 1501 AAGAAAAAGCGGAGCGGCTATTGTTGGGCGGATPATTTGCGTTAAATTGACGGCAATAT 1560
DB 1501 AAGAAAAAGCGGAGCGGCTATTGTTGGGCGGATPATTTGCGTTAAATTGACGGCAATAT 1560
QY 1561 AACGCTCAAGGTAGTGTGATATCGCTAAACCGGTGGTTTGTGGAGACATCGGGGAT 1620
DB 1561 AACGCTCAAGGTAGTGTGATATCGCTAAACCGGTGGTTTGTGGAGACATCGGGGAT 1620
QY 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAAAACAAAAAGAGTGGTTGCTAGACCCCTGAT 1680
DB 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAAAACAAAAAGAGTGGTTGCTAGACCCCTGAT 1680
QY 1681 GATGTAAACAAATTCAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGAATTC 1740
DB 1681 GATGTAAACAAATTCAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGAATTC 1740
QY 1741 CCAACAGGCGGCTGAAGCAAGGACCCCTAAAAAAATAGGAATCAAAACACGCTA 1800
DB 1741 CCAACAGGCGGCTGAAGCAAGGACCCCTAAAAAAATAGGAATCAAAACACGCTA 1800
QY 1801 ACCAATACAACTATTTCAAATATTCTGAAAAACGGCTGGCAATGAATATAACGGCATCA 1860
DB 1801 ACCAATACAACTATTTCAAATATTCTGAAAAACGGCTGGCAATGAATATAACGGCATCA 1860
QY 1861 AGAAAACTTACCGTTAAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920
DB 1861 AGAAAACTTACCGTTAAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920
QY 1921 AGTAAAGTACAGCGTGGGAGGCGTTGAGATTTGATGAGATATTACTTCTAAAGGGGA 1980
DB 1921 AGTAAAGTACAGCGTGGGAGGCGTTGAGATTTGATGAGATATTACTTCTAAAGGGGA 1980
QY 1981 AATTTAAACATTTATTCGCGGAGTGGTGTGATGTTCTATAAAATATTACGCTTTGATCAG 2040
DB 1981 AATTTAAACATTTATTCGCGGAGTGGTGTGATGTTCTATAAAATATTACGCTTTGATCAG 2040
QY 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAAGCAGC 2100
DB 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAAGCAGC 2100
QY 2101 GACGGGCAATGCTAAATTTGCGCCAGGCACTGTAAACATTACAGGAGGGA 2160
DB 2101 GACGGGCAATGCTAAATTTGCGCCAGGCACTGTAAACATTACAGGAGGGA 2160
QY 2161 GATTTACGGGCTAACAAGTATCTTTAAACGGAACGGTAAAGTCTGAATATCATTTCA 2220
DB 2161 GATTTACGGGCTAACAAGTATCTTTAAACGGAACGGTAAAGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATAATTTAACCCACAATCTTAGTGGCAACAATTACATATCTGGGAATATAACA 2280
DB 2221 TCAGTGAATAATTTAACCCACAATCTTAGTGGCAACAATTACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAACACGACCCATGATTCGACTGG 2340
DB 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAACACGACCCATGATTCGACTGG 2340
QY 2341 AACGTCAGTGTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTATTAATACATTTCA 2400
DB 2341 AACGTCAGTGTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTATTAATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAAGCTCTCGAGGGGTGAATTTTAAACGGC 2460

DB 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAAGCTCTGAGGGGTGAATTTTAAACGGC 2460
QY 2461 GTAAATGGCAACATGCTCAATCAATCTCAAGAGAGGAGGAAAGTTAAATTTCAAAATAAAA 2520
DB 2461 GTAAATGGCAACATGCTCAATCAATCTCAAGAGAGGAGGAAAGTTAAATTTCAAAATAAAA 2520
QY 2521 CCAACAGAGACATGAACACAGCAAAACCTTTACCAATTCGGTTTTAGCCAATATCACA 2580
DB 2521 CCAACAGAGACATGAACACAGCAAAACCTTTACCAATTCGGTTTTAGCCAATATCACA 2580
QY 2581 GCCACTGTGGGGCTCTGTTTTTTTTCATATATATGCCAACCATCTCGGCAGAGGGCT 2640
DB 2581 GCCACTGTGGGGCTCTGTTTTTTTTCATATATATGCCAACCATCTCGGCAGAGGGCT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGTAAATTTTACCTTAAATTCCTCAT 2700
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGTAAATTTTACCTTAAATTCCTCAT 2700
QY 2701 GTTCGCGCGGATGAGCTTTTAAAAATCAACAAAGACTTAACCATAAATGCAACCAATTC 2760
DB 2701 GTTCGCGCGGATGAGCTTTTAAAAATCAACAAAGACTTAACCATAAATGCAACCAATTC 2760
QY 2761 AATTTCAAGCTTCAGACAGAGAGAGATGTTTTTATGACGGGTACGACCAATGCGCATC 2820
DB 2761 AATTTCAAGCTTCAGACAGAGAGAGATGTTTTTATGACGGGTACGACCAATGCGCATC 2820
QY 2821 AATTTCAAGCTTCAGACATATCCATTTCTGGGGGTAAATGTCACCCCTTGGTGACAAAATCA 2880
DB 2821 AATTTCAAGCTTCAGACATATCCATTTCTGGGGGTAAATGTCACCCCTTGGTGACAAAATCA 2880
QY 2881 AGCAGCAGCATTCAGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTAGCTTAGAAGCC 2940
DB 2881 AGCAGCAGCATTCAGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTAGCTTAGAAGCC 2940
QY 2941 AATTAACGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAAACCTTGGCAGCTTGCTC 3000
DB 2941 AATTAACGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAAACCTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTTAACTTAACTGGGAAATTCAGATATTAAAGCAATCTCAGCTAT 3060
DB 3001 GTTAATGGGAGTTTAACTTAACTGGGAAATTCAGATATTAAAGCAATCTCAGCTAT 3060
QY 3061 TCAGAAAGCCCTTAAAGGAAAGACTAGAGATACCTAAATATCACCGGCAATTTT 3120
DB 3061 TCAGAAAGCCCTTAAAGGAAAGACTAGAGATACCTAAATATCACCGGCAATTTT 3120
QY 3121 ACCAATTAAGGCACTCGGAAATTAATAACAAGAGGAGTGGTAAACCTTGGCAATGTT 3180
DB 3121 ACCAATTAAGGCACTCGGAAATTAATAACAAGAGGAGTGGTAAACCTTGGCAATGTT 3180
QY 3181 ACCAATGATGTTGATTTAAACATTAACACTCAGCTAAACGCAACCAAGAGCATCATC 3240
DB 3181 ACCAATGATGTTGATTTAAACATTAACACTCAGCTAAACGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAAGGAAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
DB 3241 GCGGAGATATAATCAACAAAAGGAAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGGAAAAGAGGCAACCTCAGGATTTCTTCCGAT 3360
DB 3301 GAAATCCAAATTTGGCGGCAATATCTCGGAAAAGAGGCAACCTCAGGATTTCTTCCGAT 3360
QY 3361 AAAATTAATATCACCAACAGATAACAATAACCAAGGATTTGATGGAGGACTCTAGT 3420
DB 3361 AAAATTAATATCACCAACAGATAACAATAACCAAGGATTTGATGGAGGACTCTAGT 3420
QY 3421 TCAGATGCGACAGTAATGCCAACCTTAATACTTTAAACCAAAAGAAATGAAATTTGACAAA 3480
DB 3421 TCAGATGCGACAGTAATGCCAACCTTAATACTTTAAACCAAAAGAAATGAAATTTGACAAA 3480
QY 3481 GACCTAAGTATTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGTTTA 3540
DB 3481 GACCTAAGTATTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGTTTA 3540

Db 3481 GACCTAAGTATTTCAGGTTTCATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
Qy 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAGCCAAAACAGTAACCTTTTAAAC 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAGCCAAAACAGTAACCTTTTAAAC 3600
Qy 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTCAATGTGACACTAAATAGCAAAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTCAATGTGACACTAAATAGCAAAAGTG 3660
Qy 3661 AAAACATCTAGCAGCAATGGCGAGCTGAAAGCAATAGCAGCAACGATACCGGCTTAACT 3720
Db 3661 AAAACATCTAGCAGCAATGGCGAGCTGAAAGCAATAGCAGCAACGATACCGGCTTAACT 3720
Qy 3721 ATTACTGCAAAATCTAGCAAGTAACAAAGATATTACTTCTCAAAACAGTAATATC 3780
Db 3721 ATTACTGCAAAATCTAGCAAGTAACAAAGATATTACTTCTCAAAACAGTAATATC 3780
Qy 3781 ACCGGTGGAAAAGTTACCAACAGAGTATACAGCGGTACGATTTCCGGTAACACGGTAAGT 3900
Db 3781 ACCGGTGGAAAAGTTACCAACAGAGTATACAGCGGTACGATTTCCGGTAACACGGTAAGT 3900
Qy 3841 GCAAGTATTACAAACCAACAGGTATATACAGCGGTACGATTTCCGGTAACACGGTAAGT 3900
Db 3841 GCAAGTATTACAAACCAACAGGTATATACAGCGGTACGATTTCCGGTAACACGGTAAGT 3900
Qy 3901 GTTAGCGGACTGGTATTAAACCACTAAATCCGGTCAAAATTTGAAGCGAAATCGGGT 3960
Db 3901 GTTAGCGGACTGGTATTAAACCACTAAATCCGGTCAAAATTTGAAGCGAAATCGGGT 3960
Qy 3961 GAGCTAATGTAAAGTCAACAGGTACAAATTTGCGGGTACAAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGCTAATGTAAAGTCAACAGGTACAAATTTGCGGGTACAAATTTCCGGTAAATACGGTA 4020
Qy 4021 AATGTTACGCAACAGCTGCGGATTTAAACAGTTGGAATGGCGCAAGAAATTAATGCGACA 4080
Db 4021 AATGTTACGCAACAGCTGCGGATTTAAACAGTTGGAATGGCGCAAGAAATTAATGCGACA 4080
Qy 4081 GAAGGAGCTGACCTTTAAACCGCAACAGGAATACCTTGACTACTGAACCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGACCTTTAAACCGCAACAGGAATACCTTGACTACTGAACCGGTTCTAGC 4140
Qy 4141 ATCACTTCACTAAGGCTAGGTAGACCTCTTGCTCAGAAATGGTAGCATCGCAGGAAGC 4200
Db 4141 ATCACTTCACTAAGGCTAGGTAGACCTCTTGCTCAGAAATGGTAGCATCGCAGGAAGC 4200
Qy 4201 ATTAATGCTGCTAATGTGACATTAAATACAGGCACTTAACCCAGCTGGCAGGCTCG 4260
Db 4201 ATTAATGCTGCTAATGTGACATTAAATACAGGCACTTAACCCAGCTGGCAGGCTCG 4260
Qy 4261 GATATTAAAGCAACAGCGGACCTTGTTTATTAAACGCAAAAGATGCTAAAGCTAAATGGT 4320
Db 4261 GATATTAAAGCAACAGCGGACCTTGTTTATTAAACGCAAAAGATGCTAAAGCTAAATGGT 4320
Qy 4321 GATCATCAGGTGATAGTACAGAAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATCATCAGGTGATAGTACAGAAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Qy 4381 ACTGGGCAACCTCAACAGGTGCAATATCACTGGGATTTAAACAGTAATTAATGGGTTA 4440
Db 4381 ACTGGGCAACCTCAACAGGTGCAATATCACTGGGATTTAAACAGTAATTAATGGGTTA 4440
Qy 4441 AATATCAATTCGAAGATGGTAGAAACACTGTGCGCTTAAAGAGCAAGGAATTTGAGGTG 4500
Db 4441 AATATCAATTCGAAGATGGTAGAAACACTGTGCGCTTAAAGAGCAAGGAATTTGAGGTG 4500
Qy 4501 AAATATATCCAGCGAGGTGTAGCAAGTGTAGAAGAAATTAATGAAGCGAAACGGTCCCT 4560
Db 4501 AAATATATCCAGCGAGGTGTAGCAAGTGTAGAAGAAATTAATGAAGCGAAACGGTCCCT 4560
Qy 4561 GAAAAAGTAAAGATTATTCTGATGAGAAAGAAACATTAGCTAAACCTTTGGTGTAAAGT 4620
Db 4561 GAAAAAGTAAAGATTATTCTGATGAGAAAGAAACATTAGCTAAACCTTTGGTGTAAAGT 4620

Qy 4621 GCTGTAGCTTTTGTGAGCCAAATTAATACAAATTAACAGTCAATACACAAAATGAATTTACA 4680
Db 4621 GCTGTAGCTTTTGTGAGCCAAATTAATACAAATTAACAGTCAATACACAAAATGAATTTACA 4680
Qy 4681 ACCAGACCGTCAAGTCAAGTCAATTTCTGAAGGTAAAGCGTGTCTCAAGTGGTAAT 4740
Db 4681 ACCAGACCGTCAAGTCAAGTCAATTTCTGAAGGTAAAGCGTGTCTCAAGTGGTAAT 4740
Qy 4741 GGCACACGAGTATGACCAATGTTGCTGACGATGGACGCGTGTAGTCAATTAATGACAAG 4800
Db 4741 GGCACACGAGTATGACCAATGTTGCTGACGATGGACGCGTGTAGTCAATTAATGACAAG 4800
Qy 4801 GTAGATTTCATCCCTGCAATGAAGTCAATTTATTTTCGTATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTCATCCCTGCAATGAAGTCAATTTATTTTCGTATTTACTGTGGGTTAAA 4860
Qy 4861 GTTCAGTACGGGCTTTACCCATCTTGTAAGAAATTTACGGAGATACAAATAAGTATTTT 4920
Db 4861 GTTCAGTACGGGCTTTACCCATCTTGTAAGAAATTTACGGAGATACAAATAAGTATTTT 4920
Qy 4921 AACAGTTATTATTATG 4937
Db 4921 AACAGTTATTATTATG 4937

RESULT 5
LOCUS I35771 4937 bp DNA linear PAT 13-MAY-1997
DEFINITION Sequence 3 from patent US 5603938.
ACCESSION I35771
VERSION I35771.1 GI:2086995
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4937)
AUTHORS Barenkamp,S.J.
TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 5603938-A 3 18-FEB-1997;
FEATURES
Location/Qualifiers
1. 4937
Source
BASE COUNT 1729 a 948 c 1009 g 1251 t
ORIGIN
Query Match 100.0%; Score 4937; DB 6; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAAATATCAAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 60
Db 1 TAAATATCAAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 60
Qy 61 CACCTTTTGGCAGTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120
Db 61 CACCTTTTGGCAGTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120
Qy 121 AATGGTATATCTTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
Db 121 AATGGTATATCTTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
Qy 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
Db 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
Qy 241 CACATGAATGATCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 241 CACATGAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Qy 301 GAACGCAATGATAAGTAAATTAATTTGTTCACTACCTTAGGAGAAATATGACAAG 360
Db 301 GAACGCAATGATAAGTAAATTAATTTGTTCACTACCTTAGGAGAAATATGACAAG 360

QY 361 ATATATCGTCTCAAAATTCAGCAACCCCTGAATGCTTTGGTTGCTGTCTGAATGGCA 420
DB 361 ATATATCGTCTCAAAATTCAGCAACCCCTGAATGCTTTGGTTGCTGTCTGAATGGCA 420
QY 421 CGGGGTGTGACCATTTCCACAGAAAAGGCTTCGGCTATGTTTACATCTTTAGGTGAAC 480
DB 421 CGGGGTGTGACCATTTCCACAGAAAAGGCTTCGGCTATGTTTACATCTTTAGGTGAAC 480
QY 481 CACTTAGCGTTAAAGCACTTTCCGCTATGTTTACTATCTTTAGGTGTAACATCTTCCA 540
DB 481 CACTTAGCGTTAAAGCACTTTCCGCTATGTTTACTATCTTTAGGTGTAACATCTTCCA 540
QY 541 CAATCTGTTTTAGCAAGCGGCTTTACAAGGAATGGATAGTACACGGCACAGCCACTATG 600
DB 541 CAATCTGTTTTAGCAAGCGGCTTTACAAGGAATGGATAGTACACGGCACAGCCACTATG 600
QY 601 CAAGTAGATGGTAATAAACCATTATCCGCAACAGCTGTTGACGCTATCATTTAATTTGAAA 660
DB 601 CAAGTAGATGGTAATAAACCATTATCCGCAACAGCTGTTGACGCTATCATTTAATTTGAAA 660
QY 661 CAATTTAATCAATCAAAATGGTGGAGTTTACAGAAAACCAACCACTCCGCC 720
DB 661 CAATTTAATCAATCAAAATGGTGGAGTTTACAGAAAACCAACCACTCCGCC 720
QY 721 GTATTCAACCGGTTACATCTTAACCAATCTCCCAATTTAAAGGATTTTATGATTTAAC 780
DB 721 GTATTCAACCGGTTACATCTTAACCAATCTCCCAATTTAAAGGATTTTATGATTTAAC 780
QY 781 GGCAAGTCTTTTTAATCAACCAATGGTATCAATAGTAAAGCGCAATTTATTAAC 840
DB 781 GGCAAGTCTTTTTAATCAACCAATGGTATCAATAGTAAAGCGCAATTTATTAAC 840
QY 841 ACTAATGCTTTACGGCTTCTACGCTAGACATTTCTTAACGAAAACATCAAGCGCGTAAT 900
DB 841 ACTAATGCTTTACGGCTTCTACGCTAGACATTTCTTAACGAAAACATCAAGCGCGTAAT 900
QY 901 TTACCTTCGAGCAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCAGCGTTTAAAT 960
DB 901 TTACCTTCGAGCAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCAGCGTTTAAAT 960
QY 961 ACTGTCGTAAGACGGGAGTGAATCTTAATGTTGGTGGCAAGTGAAGGAGGTTG 1020
DB 961 ACTGTCGTAAGACGGGAGTGAATCTTAATGTTGGTGGCAAGTGAAGGAGGTTG 1020
QY 1021 ATTAGCGTAAATGTTGGCAGCATTTCTTACTCGCGGCAAGAAAATCACCATCAGCGAT 1080
DB 1021 ATTAGCGTAAATGTTGGCAGCATTTCTTACTCGCGGCAAGAAAATCACCATCAGCGAT 1080
QY 1081 ATAATAACCCCAACATTTACTTACAGCATTTGCCGCGCTGAAATGAAGCGGTCAATCTG 1140
DB 1081 ATAATAACCCCAACATTTACTTACAGCATTTGCCGCGCTGAAATGAAGCGGTCAATCTG 1140
QY 1141 GGCGATATTTTCCAAAGCGGTTAAATTAATGTCGCTGCTGCGCATTTTCGAAACCAA 1200
DB 1141 GGCGATATTTTCCAAAGCGGTTAAATTAATGTCGCTGCTGCGCATTTTCGAAACCAA 1200
QY 1201 GGTAACTTTCTGCTGATTTCTGAAGCAAGATTAAGCGGCAATTTGTTCTTCGCGC 1260
DB 1201 GGTAACTTTCTGCTGATTTCTGAAGCAAGATTAAGCGGCAATTTGTTCTTCGCGC 1260
QY 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320
DB 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320
QY 1321 GGCAAGCTGATGATTACAGCGGATAAGTCAATTAACACAGTGCAGTTATCGACCTT 1380
DB 1321 GGCAAGCTGATGATTACAGCGGATAAGTCAATTAACACAGTGCAGTTATCGACCTT 1380
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGGCAAGGTAATAAAC 1440
DB 1381 TCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGGCAAGGTAATAAAC 1440

QY 1441 GGCATTCAAATTAGCAAAAGAAACCTCTTTGAAAAGAGCTCAACCATCAATTTATCAGGC 1500
DB 1441 GGCATTCAAATTAGCAAAAGAAACCTCTTTGAAAAGAGCTCAACCATCAATTTATCAGGC 1500
QY 1501 AAAGAAAAGCGGACGGCTATTGTTGGGGCGATATTGGGTTAAATTTAGCGGCAATTT 1560
DB 1501 AAAGAAAAGCGGACGGCTATTGTTGGGGCGATATTGGGTTAAATTTAGCGGCAATTT 1560
QY 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGGTTTGTGGAGACATCGGGCAT 1620
DB 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGGTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTTATCCATTGACGCAATTTGTTTAAACAAAAGAGTGGTTGCTAGACCTGAT 1680
DB 1621 TATTTATCCATTGACGCAATTTGTTTAAACAAAAGAGTGGTTGCTAGACCTGAT 1680
QY 1681 GATCTAACAATTTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATAAATGATGATTC 1740
DB 1681 GATCTAACAATTTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATAAATGATGATTC 1740
QY 1741 CCAACAGCGCGGTGAAGCAAGGACCCCTTAAAATAATAGCGAACTCAAAACACGCTA 1800
DB 1741 CCAACAGCGCGGTGAAGCAAGGACCCCTTAAAATAATAGCGAACTCAAAACACGCTA 1800
QY 1801 ACCAATAACAATTTTCAAAATTTATCTGAAAACCGCTTGAACAATGAATATAAGCGCATCA 1860
DB 1801 ACCAATAACAATTTTCAAAATTTATCTGAAAACCGCTTGAACAATGAATATAAGCGCATCA 1860
QY 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCACCTTAATTTCCAT 1920
DB 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCACCTTAATTTCCAT 1920
QY 1921 AGTAAAGGTGAGCGTGGCGAGCGTTTCAAGATTTAGGAGATTAATCTTAAAGCGGA 1980
DB 1921 AGTAAAGGTGAGCGTGGCGAGCGTTTCAAGATTTAGGAGATTAATCTTAAAGCGGA 1980
QY 1981 AATTTAACCAATTTATCTGCGGATGGTTGATTTTATATAAATAATTTACGCTTGATCAG 2040
DB 1981 AATTTAACCAATTTATCTGCGGATGGTTGATTTTATATAAATAATTTACGCTTGATCAG 2040
QY 2041 GGTTTTTTAAATTTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAAGCACGC 2100
DB 2041 GGTTTTTTAAATTTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAAGCACGC 2100
QY 2101 GAGCGGCAAAATGCTAAAATTTGTCGCCAGGCACTGTAAACATTACAGGAGAGGAAAA 2160
DB 2101 GAGCGGCAAAATGCTAAAATTTGTCGCCAGGCACTGTAAACATTACAGGAGAGGAAAA 2160
QY 2161 GATTTGAGGCTTAAACACGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2220
DB 2161 GATTTGAGGCTTAAACACGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTTAACCCACATCTTACTGGCACAATTAACATATCTGGGAATATAACA 2280
DB 2221 TCAGTGAATTAATTTAACCCACATCTTACTGGCACAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACAGCATGATTCGCACTGG 2340
DB 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACAGCATGATTCGCACTGG 2340
QY 2341 AACGTCAGTGTCTTAACTAGACACGCGCAATTTTACCTTTTAAATATCATTTCA 2400
DB 2341 AACGTCAGTGTCTTAACTAGACACGCGCAATTTTACCTTTTAAATATCATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGACGGGTGAATTTTAAACGC 2460
DB 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGACGGGTGAATTTTAAACGC 2460
QY 2461 GTAAATGCAACATCTCATTTCAATCTCAAGAGAGGCAAGTTAATTTCAAAATTAATA 2520
DB 2461 GTAAATGCAACATCTCATTTCAATCTCAAGAGAGGCAAGTTAATTTCAAAATTAATA 2520
QY 2521 CCAACAGAGACATGAACACAAAGCAACCTTTTACCAATTCGTTTTTAGCCCAATATCACA 2580

|||||
Db 2521 CCAACGAGACATGAAACAGCAACCTTTACCAATTCGGTTTATGCGCAATATACACA 2580
QY 2581 GCCACTGGTGGGGCTCTCTTTTGGATATATATGCGCAACCATCTGGCAGAGGGCT 2640
Db 2581 GCCACTGGTGGGGCTCTCTTTTGGATATATATGCGCAACCATCTGGCAGAGGGCT 2640
QY 2641 GAGTTAAAATGAGTGAATATATCTCTACGGCGCTTAATTTTACCTTAATTTCCCAT 2700
Db 2641 GAGTTAAAATGAGTGAATATATCTCTACGGCGCTTAATTTTACCTTAATTTCCCAT 2700
QY 2701 GTTCGGGGGATGACGCTTTTAAATCAACAAGAGCTTAACCAATAATGCAACCAATTC 2760
Db 2701 GTTCGGGGGATGACGCTTTTAAATCAACAAGAGCTTAACCAATAATGCAACCAATTC 2760
QY 2761 AATTTTCAGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCCATC 2820
Db 2761 AATTTTCAGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCCATC 2820
QY 2821 AATTTCAACCTTACACATATCCATCTTGGCGGTAAATGTCACCCCTTGGTGGCAAAATCTCA 2880
Db 2821 AATTTCAACCTTACACATATCCATCTTGGCGGTAAATGTCACCCCTTGGTGGCAAAATCTCA 2880
QY 2881 AGCAGAGCATTTACGGGGATATATCTATCGAGAAAGCAGCAATGTTACGCTAGAGCC 2940
Db 2881 AGCAGAGCATTTACGGGGATATATCTATCGAGAAAGCAGCAATGTTACGCTAGAGCC 2940
QY 2941 AATAAGCCCTTAATCAGCAAAACATAAAGGATAGAGTTATATAAACTTGGCAGCTTGCTC 3000
Db 2941 AATAAGCCCTTAATCAGCAAAACATAAAGGATAGAGTTATATAAACTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTTAACTTGGCGAAATGCGAGATATTAAGAGCAATCTCACTAT 3060
Db 3001 GTTAATGGGAGTTTAACTTGGCGAAATGCGAGATATTAAGAGCAATCTCACTAT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGAAAGACTAGAGATACCCCTAAATATACCGCGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGAAAGACTAGAGATACCCCTAAATATACCGCGCAATTTT 3120
QY 3121 ACCAATATGCGCTGCGGAAATTAATATACCAAGAGTGTGTAACACTTGGCAATGTT 3180
Db 3121 ACCAATATGCGCTGCGGAAATTAATATACCAAGAGTGTGTAACACTTGGCAATGTT 3180
QY 3181 ACCAATGATGTTTAAACATTAACCTACGCTTAACGCAACCAAGAGCATCATC 3240
Db 3181 ACCAATGATGTTTAAACATTAACCTACGCTTAACGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATATCAACAAAAGAGAGCTTAATATTAACAGTACAGCATTAATGATGCT 3300
Db 3241 GCGGAGATATATCAACAAAAGAGAGCTTAATATTAACAGTACAGCATTAATGATGCT 3300
QY 3301 GAAATCCAAATGCGGCAATATCTCGCAAAAAGAGGCAACCTCAGCATTTCTCCGAT 3360
Db 3301 GAAATCCAAATGCGGCAATATCTCGCAAAAAGAGGCAACCTCAGCATTTCTCCGAT 3360
QY 3361 AAAATTAATATCAACCAACAGATTAACATAAAGAGGTATTTGATGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATATCAACCAACAGATTAACATAAAGAGGTATTTGATGAGAGGACTCTAGT 3420
QY 3421 TCAGATGCGCAAGTAAATGCGCAACCTAACTATTAACCAAGAGTAAATGCAAGAA 3480
Db 3421 TCAGATGCGCAAGTAAATGCGCAACCTAACTATTAACCAAGAGTAAATGCAAGAA 3480
QY 3481 GACCTAAGTATTCAGGTTTCAATTAAGCAGAGATTAACGCAAGAGTGTAGAGATTTA 3540
Db 3481 GACCTAAGTATTCAGGTTTCAATTAAGCAGAGATTAACGCAAGAGTGTAGAGATTTA 3540
QY 3541 ACTATTGCAACAGTAAATGAGCGTAAACAGCGGTGCGGAGCCCAACAGTAACCTTTAAC 3600
Db 3541 ACTATTGCAACAGTAAATGAGCGTAAACAGCGGTGCGGAGCCCAACAGTAACCTTTAAC 3600
QY 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTCAACATGTGACACTAAATAGCAAGTG 3660
|||||

Db 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTCAACATGTGACACTAAATAGCAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCGCAACAGATACCGGCTTAAC 3720
Db 3661 AAAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCGCAACAGATACCGGCTTAAC 3720
QY 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTAATCTCTCTCAAAACAGTAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTAATCTCTCTCAAAACAGTAATATC 3780
QY 3781 ACCGGTCTGGAAGAGTTTACCACACAGCAGCTCGACCATTAACGCAACAAATGGCAAA 3840
Db 3781 ACCGGTCTGGAAGAGTTTACCACACAGCAGCTCGACCATTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACAAACCAAGGTGATATCAGCGGTAGCATTTCCGGTAACACGCTAAGT 3900
Db 3841 GCAAGTATTACAAACCAAGGTGATATCAGCGGTAGCATTTCCGGTAACACGCTAAGT 3900
QY 3901 GTTAGCGGACTGGTGTATTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960
Db 3901 GTTAGCGGACTGGTGTATTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960
QY 3961 GAGGCTAATGTAACAAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAATACGGTA 4020
Db 3961 GAGGCTAATGTAACAAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAATACGGTA 4020
QY 4021 AATGTTACGGCAACGCTGGCGATTTAACAGTTGGGAATGGCGCAAAATTAATCGGACA 4080
Db 4021 AATGTTACGGCAACGCTGGCGATTTAACAGTTGGGAATGGCGCAAAATTAATCGGACA 4080
QY 4081 GAAGGAGCTGAACCTTTAACCGCAACAGGATACCTTTGACTACTGAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGAACCTTTAACCGCAACAGGATACCTTTGACTACTGAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTTGGCTCAGAAATGGTAGATCCGCAAGC 4200
Db 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTTGGCTCAGAAATGGTAGATCCGCAAGC 4200
QY 4201 ATTAATGCTGTAAATGTGACATTAATACTACAGGCACTTAACCAAGCTGCGAGGCTG 4260
Db 4201 ATTAATGCTGTAAATGTGACATTAATACTACAGGCACTTAACCAAGCTGCGAGGCTG 4260
QY 4261 GATATTAAAGCAACAGCGGCACTTTGGTTATTAAAGCAAAAAGATGCTAAGCTAATGTT 4320
Db 4261 GATATTAAAGCAACAGCGGCACTTTGGTTATTAAAGCAAAAAGATGCTAAGCTAATGTT 4320
QY 4321 GATGCAATCAGGTGATAGTACAGAAAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4380
Db 4321 GATGCAATCAGGTGATAGTACAGAAAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4380
QY 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGGTTA 4440
Db 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGGTTA 4440
QY 4441 AATATCATTTGCAAGATGTTAGAACACTGTCCGCTTAAGAGCAAGCAAAATTCAGGTTG 4500
Db 4441 AATATCATTTGCAAGATGTTAGAACACTGTCCGCTTAAGAGCAAGCAAAATTCAGGTTG 4500
QY 4501 AATATATATCCAGCCAGGTGTAGCAAGTGTAGAAGAGTAAATTTGAAGCGCAAAATTCAG 4560
Db 4501 AATATATATCCAGCCAGGTGTAGCAAGTGTAGAAGAGTAAATTTGAAGCGCAAAATTCAG 4560
QY 4561 GAAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTAAGCTTGGTGAAGT 4620
Db 4561 GAAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTAAGCTTGGTGAAGT 4620
QY 4621 GCTGTACGTTTCTGAGCCAAATTAATACAAATTAACAGTCAATACACAAATTAATTAAC 4680
Db 4621 GCTGTACGTTTCTGAGCCAAATTAATACAAATTAACAGTCAATACACAAATTAATTAAC 4680
QY 4681 ACCAGACGCTCAAGTCAAGTGAATTTTCGAAAGTAAAGCGGTTTCTCAAGTGTAAAT 4740
Db 4681 ACCAGACGCTCAAGTCAAGTGAATTTTCGAAAGTAAAGCGGTTTCTCAAGTGTAAAT 4740

```

QY 4741 GCGCAGGAGTATGTACCAATGTGCTGACGATGGACGCCGTAGTCAAGTAAATTGACAAG 4800
|
|
|
Db 4741 GCGCAGGAGTATGTACCAATGTGCTGACGATGGACGCCGTAGTCAAGTAAATTGACAAG 4800
|
|
|
QY 4801 GTAGATTTCATCTGCAATGAAGTCAATTTATTTTCGTAATTTACTGTGTGGTTAAA 4860
|
|
|
Db 4801 GTAGATTTCATCTGCAATGAAGTCAATTTATTTTCGTAATTTACTGTGTGGTTAAA 4860
|
|
|
QY 4861 GTTACGTACGGGCTTTACCCATCTTTGTAATAAATACGGAGAAATACAATAAAGTATTTT 4920
|
|
|
Db 4861 GTTACGTACGGGCTTTACCCATCTTTGTAATAAATACGGAGAAATACAATAAAGTATTTT 4920
|
|
|
QY 4921 AACAGTTTATTATTATG 4937
|
|
|
Db 4921 AACAGTTTATTATTATG 4937
|
|
|
RESULT 6
BD010356
LOCUS
DEFINITION High molecular weight surface protein of non-typeable haemophils.
ACCESSION BD010356
VERSION BD010356.1 GI:18638729
KEYWORDS JP 2001503602-A/2.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 4937)
AUTHORS Barenkamp,S.J.
TITLE High molecular weight surface protein of non-typeable haemophils
JOURNAL Patent: JP 2001503602-A 2 21-MAR-2001;
COMMENT SENSITIVE UNIV, WASHINGTON UNIV
OS Haemophils
PN JP 2001503602-A/2
PD 21-MAR-2001
PE 01-APR-1997 JP 1997535346
PR 01-APR-1996 US 08/617697
PI STEPHEN J BARENKAMP
PC C07H21/02,C07H21/04,C12P21/06,A61K39/102
CC
FH
FT
FT
FEATURES
source Location/Qualifiers
1..4937
/organism="Haemophils"
Location/Qualifiers
1..4937
/organism="unclassified"
/db_xref="taxon:32644"
BASE COUNT 1728 a 949 c 1009 g 1251 t
ORIGIN
Query Match 100.0%; Score 4935.4; DB 6; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4936; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TAAATATACAGATATAATAAATAAATCAAGATTTTGTGATGACAAACAACAATTACAA 60
|
|
|
Db 1 TAAATATACAGATATAATAAATAAATCAAGATTTTGTGATGACAAACAACAATTACAA 60
|
|
|
QY 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
|
|
|
Db 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
|
|
|
QY 121 AATGGTATATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
|
|
|
Db 121 AATGGTATATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
|
|
|
QY 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
|
|
|
Db 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
|
|
|
QY 241 CACATGAATGATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
|
|
|

```

```

Db 241 CACATGAATGATCAACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
|
|
|
QY 301 CAACGCAATGATTAAGTAAATTTAATTGTTCAACTAACCTTTAGGAGAAAATATGAAACAAG 360
|
|
|
Db 301 CAACGCAATGATTAAGTAAATTTAATTGTTCAACTAACCTTTAGGAGAAAATATGAAACAAG 360
|
|
|
QY 361 ATATATGCTCTCAAAATTCAGCAAAACGCTGAATGCTTTGGTTGCTGCTGCTGCTGCTGCTG 420
|
|
|
Db 361 ATATATGCTCTCAAAATTCAGCAAAACGCTGAATGCTTTGGTTGCTGCTGCTGCTGCTGCTG 420
|
|
|
QY 421 CGGGGTTGTGACCAATTCACAGAAAAGGCTTCGCGCTATGTTACTATCTTTTAGGTGTAAAC 480
|
|
|
Db 421 CGGGGTTGTGACCAATTCACAGAAAAGGCTTCGCGCTATGTTACTATCTTTTAGGTGTAAAC 480
|
|
|
QY 481 CACTTAGCGTTAAAGCCACATTTCCGCTATGTTACTATCTTTTAGGTGTAAACATCTATTCCA 540
|
|
|
Db 481 CACTTAGCGTTAAAGCCACATTTCCGCTATGTTACTATCTTTTAGGTGTAAACATCTATTCCA 540
|
|
|
QY 541 CAATCTCTTTTAGCAAGCGGCTTACAAAGGAATGGATGTACACGGCACAGCCACTATG 600
|
|
|
Db 541 CAATCTCTTTTAGCAAGCGGCTTACAAAGGAATGGATGTACACGGCACAGCCACTATG 600
|
|
|
QY 601 CAAGTAGATGTTAATAAACCAATTCGCCAACAGTGTTCACGCTATCTATTAATTGAAA 660
|
|
|
Db 601 CAAGTAGATGTTAATAAACCAATTCGCCAACAGTGTTCACGCTATCTATTAATTGAAA 660
|
|
|
QY 661 CAATTTAACATCGACCAAAATGAATGGTGCAGTGTTCACAAAGAAACAACAACACTCCGCC 720
|
|
|
Db 661 CAATTTAACATCGACCAAAATGAATGGTGCAGTGTTCACAAAGAAACAACAACACTCCGCC 720
|
|
|
QY 721 GTATTCAACCGTGTTCATCTAACCAAAATCTCCCAATTTAAAGGGGATTTTAGATTCTAAC 780
|
|
|
Db 721 GTATTCAACCGTGTTCATCTAACCAAAATCTCCCAATTTAAAGGGGATTTTAGATTCTAAC 780
|
|
|
QY 781 GGACAAAGTCTTTTAAATCAACCCAAATGGTATCACAATAGGTAAAGACGAATTTATTAAAC 840
|
|
|
Db 781 GGACAAAGTCTTTTAAATCAACCCAAATGGTATCACAATAGGTAAAGACGAATTTATTAAAC 840
|
|
|
QY 841 ACTAATGGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGGCGGTAAT 900
|
|
|
Db 841 ACTAATGGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGGCGGTAAT 900
|
|
|
QY 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
|
|
|
Db 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
|
|
|
QY 961 ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTGTTGCAAAAGTGAAGAACAGGAGGTG 1020
|
|
|
Db 961 ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTGTTGCAAAAGTGAAGAACAGGAGGTG 1020
|
|
|
QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTTTACTCGAGGGCAAAAATACCATCAGCGAT 1080
|
|
|
Db 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTTTACTCGAGGGCAAAAATACCATCAGCGAT 1080
|
|
|
QY 1081 ATAATAAACCACCAATTTACTTTACAGCATTTGCGGCGCTGAAATGAAGCGGTCAATCTG 1140
|
|
|
Db 1081 ATAATAAACCACCAATTTACTTTACAGCATTTGCGGCGCTGAAATGAAGCGGTCAATCTG 1140
|
|
|
QY 1141 GCGCATATTTTGGCAAAAGCGGTAAACATTAATGTCGCTGCTGCCACTATTTCGAACCAA 1200
|
|
|
Db 1141 GCGCATATTTTGGCAAAAGCGGTAAACATTAATGTCGCTGCTGCCACTATTTCGAACCAA 1200
|
|
|
QY 1201 GGTAAACTTTCGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTGTTCTTTCCGCC 1260
|
|
|
Db 1201 GGTAAACTTTCGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTGTTCTTTCCGCC 1260
|
|
|
QY 1261 AAAGAGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
|
|
|
Db 1261 AAAGAGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
|
|
|
QY 1321 GGCAGCTGATGATTACAGGCGGATAAAGTCAATTTAAAAACAGGTGACGTTATCGACCTT 1380
|
|
|
Db 1321 GGCAGCTGATGATTACAGGCGGATAAAGTCAATTTAAAAACAGGTGACGTTATCGACCTT 1380
|
|
|

```

Qy 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGAGAGCGCGCGGGAAGGTAAAGAC 1440
 Db 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGAGAGCGCGCGGGAAGGTAAAGAC 1440
 Qy 1441 GGCATTCAATTTAGCAAGAAACCTCTTTAGAAAAGAGGTCAACCATCATATATCAGGC 1500
 Db 1441 GGCATTCAATTTAGCAAGAAACCTCTTTAGAAAAGAGGTCAACCATCATATATCAGGC 1500
 Qy 1501 AAAGAAAAGCGGAGCGCTATTGTGTGGGCGGATATTGCGTTAATTGACGGCAATATT 1560
 Db 1501 AAAGAAAAGCGGAGCGCTATTGTGTGGGCGGATATTGCGTTAATTGACGGCAATATT 1560
 Qy 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGGTTTGTGGAGACATCGGGGCAT 1620
 Db 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGGTTTGTGGAGACATCGGGGCAT 1620
 Qy 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGTGTGCTAGACCTGAT 1680
 Db 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGTGTGCTAGACCTGAT 1680
 Qy 1681 GATGTACAAATTGAAGCCGAGACCCCTTTCGCAATAATACCGGTATTAATGATGATTC 1740
 Db 1681 GATGTACAAATTGAAGCCGAGACCCCTTTCGCAATAATACCGGTATTAATGATGATTC 1740
 Qy 1741 CCAACAGGCGCGGTGAAGCAAGCGCCCTTAAAAAAGTGAAGTCAAAACAGCGTA 1800
 Db 1741 CCAACAGGCGCGGTGAAGCAAGCGCCCTTAAAAAAGTGAAGTCAAAACAGCGTA 1800
 Qy 1801 ACCAATCAACATTTTCAAAATTTATCTGAAAAGCGCTTGAACAATGAATATAAGCGCATCA 1860
 Db 1801 ACCAATCAACATTTTCAAAATTTATCTGAAAAGCGCTTGAACAATGAATATAAGCGCATCA 1860
 Qy 1861 AGAAACCTTACCGTTAATAGTCAATCAACATCGGAAGCAACTCCCACTTAATTTCTCCAT 1920
 Db 1861 AGAAACCTTACCGTTAATAGTCAATCAACATCGGAAGCAACTCCCACTTAATTTCTCCAT 1920
 Qy 1921 AGTAAAGTCAAGCGTGGGAGGCGTTCAGATTGATCGAGATATTACTTCTAAAGCGGA 1980
 Db 1921 AGTAAAGTCAAGCGTGGGAGGCGTTCAGATTGATCGAGATATTACTTCTAAAGCGGA 1980
 Qy 1981 AATTTAACCATTATCTGGCGATGGTTGATGTTTCATATAAATATTACGCTTGATCAG 2040
 Db 1981 AATTTAACCATTATCTGGCGATGGTTGATGTTTCATATAAATATTACGCTTGATCAG 2040
 Qy 2041 GGTTTTTTAATATTACCGCGCTTCGTTAGAGTGAAGTGAATACAAAGCAGC 2100
 Db 2041 GGTTTTTTAATATTACCGCGCTTCGTTAGAGTGAAGTGAATACAAAGCAGC 2100
 Qy 2101 GACGCGCAATGCTAAATTTGTCGCGGAGGCTACTGTACCAATTACAGGAGGGAATA 2160
 Db 2101 GACGCGCAATGCTAAATTTGTCGCGGAGGCTACTGTACCAATTACAGGAGGGAATA 2160
 Qy 2161 GATTCAGGGGTAACAGCTATCTTTAAACGGAACGGGTAAAGTCTGAATATCATTTCA 2220
 Db 2161 GATTCAGGGGTAACAGCTATCTTTAAACGGAACGGGTAAAGTCTGAATATCATTTCA 2220
 Qy 2221 TCAGTGAATAATTAAACCCCAATCTTAGTGGCACAATTAAACATATCTGGGAATATAACA 2280
 Db 2221 TCAGTGAATAATTAAACCCCAATCTTAGTGGCACAATTAAACATATCTGGGAATATAACA 2280
 Qy 2281 ATTAACCAACTACGAGAAAGAACACCTCGATTGTCGCAACACCGCATGATTCGCACTGG 2340
 Db 2281 ATTAACCAACTACGAGAAAGAACACCTCGATTGTCGCAACACCGCATGATTCGCACTGG 2340
 Qy 2341 AACGTCAGTGTCTTAATCTAGACAGCGCGCAATTTTACCTTTTAAATACATTTCA 2400
 Db 2341 AACGTCAGTGTCTTAATCTAGACAGCGCGCAATTTTACCTTTTAAATACATTTCA 2400
 Qy 2401 AGCAATAGCAAGGGTTAAACACACAGATATAGACGCTTCGAGGGGTGAATTTTAAACGGC 2460
 Db 2401 AGCAATAGCAAGGGTTAAACACACAGATATAGACGCTTCGAGGGGTGAATTTTAAACGGC 2460

Qy 2461 GTAAATGCAACATGCTCAATCTCAAGAGGAGCGAAAGTTAATTTCAATTTAAA 2520
 Db 2461 GTAAATGCAACATGCTCAATCTCAAGAGGAGCGAAAGTTAATTTCAATTTAAA 2520
 Qy 2521 CCAACAGAGACATGAACACAGCAACCTTTTACCAATTCGGTTTTTACCAATATATCA 2580
 Db 2521 CCAACAGAGACATGAACACAGCAACCTTTTACCAATTCGGTTTTTACCAATATATCA 2580
 Qy 2581 GCCACTGTGGGGCTCTGTTTTTATATATATGCAACCAATCTCTGGCAGAGGGCT 2640
 Db 2581 GCCACTGTGGGGCTCTGTTTTTATATATATGCAACCAATCTCTGGCAGAGGGCT 2640
 Qy 2641 GAGTTTAAAAATGAGTGAATTAATATCTCAACGGCGCTTAATTTTACCTTTAAATTTCCCAT 2700
 Db 2641 GAGTTTAAAAATGAGTGAATTAATATCTCAACGGCGCTTAATTTTACCTTTAAATTTCCCAT 2700
 Qy 2701 GTTTCGGCGGATGACGCTTTTAAAAATCAACAAAGACTTTAAACCATAAATGCAACCAATTTCA 2760
 Db 2701 GTTTCGGCGGATGACGCTTTTAAAAATCAACAAAGACTTTAAACCATAAATGCAACCAATTTCA 2760
 Qy 2761 AATTTACGCTCAGACAGACGAAAGATGATTTTATGAGGGTACGCGCAATGCCATC 2820
 Db 2761 AATTTACGCTCAGACAGACGAAAGATGATTTTATGAGGGTACGCGCAATGCCATC 2820
 Qy 2821 AATTTCAACCTCAACATATCTGCGGCTGATGTCACCGCTTGGTGGCAAACTCA 2880
 Db 2821 AATTTCAACCTCAACATATCTGCGGCTGATGTCACCGCTTGGTGGCAAACTCA 2880
 Qy 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGAGCAAAATGTTACGGTAGAAGC 2940
 Db 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGAGCAAAATGTTACGGTAGAAGC 2940
 Qy 2941 AATAAGCGCCCTTAATCAGCAAAACATAGGGATAGAGTTATAAATACCTGGCAGTTGCTC 3000
 Db 2941 AATAAGCGCCCTTAATCAGCAAAACATAGGGATAGAGTTATAAATACCTGGCAGTTGCTC 3000
 Qy 3001 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAAATGAGATATTAAAGGCAATCTCACTATT 3060
 Db 3001 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAAATGAGATATTAAAGGCAATCTCACTATT 3060
 Qy 3061 TCAGAAAGCGCCCTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120
 Db 3061 TCAGAAAGCGCCCTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120
 Qy 3121 ACCAATTAATGGCATCTCCGAAATTAATATAACACAGGAGTGTAAACCTTGGCAATGTT 3180
 Db 3121 ACCAATTAATGGCATCTCCGAAATTAATATAACACAGGAGTGTAAACCTTGGCAATGTT 3180
 Qy 3181 ACCAATTAATGGCATTTAAACATTTAAACATTTAAACGCAACCAAGAGCAATCATC 3240
 Db 3181 ACCAATTAATGGCATTTAAACATTTAAACATTTAAACGCAACCAAGAGCAATCATC 3240
 Qy 3241 GCGGAGATATATCAACAAAAGAGAGCTTTAAATATTACAGACAGTAAATGATGCT 3300
 Db 3241 GCGGAGATATATCAACAAAAGAGAGCTTTAAATATTACAGACAGTAAATGATGCT 3300
 Qy 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTACGATTTCTCCGAT 3360
 Db 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTACGATTTCTCCGAT 3360
 Qy 3361 AAAATTAATATCAACAAACAGATTAACATTAACAAAAGGGTATTGATGGAGAGGACTTAGT 3420
 Db 3361 AAAATTAATATCAACAAACAGATTAACATTAACAAAAGGGTATTGATGGAGAGGACTTAGT 3420
 Qy 3421 TCAGATGCGCAACAGTAAATGCAACCTTAATTTAAAAACCAAGAAATTTGAAATTTGACAGAA 3480
 Db 3421 TCAGATGCGCAACAGTAAATGCAACCTTAATTTAAAAACCAAGAAATTTGAAATTTGACAGAA 3480
 Qy 3481 GACCTAGTATTTCAGGTTTCAATTAAGCAGATTTACGCCAAAGATGTTAGAGATTTA 3540
 Db 3481 GACCTAGTATTTCAGGTTTCAATTAAGCAGATTTACGCCAAAGATGTTAGAGATTTA 3540
 Qy 3541 ACTATTGGCAACAGTAAATGACGGTAACAGCGGTGCCGAAAGCCCAACAGTAACTTTTAAAC 3600

QY 961 ACTGTCGTAAGACGCGAGTGTAAATCTTATTGGTGGCAAGTGAAAAACGAGGGTGTG 1020
DB 961 ACTGTCGTAAGACGCGAGTGTAAATCTTATTGGTGGCAAGTGAAAAACGAGGGTGTG 1020
QY 1021 ATTAGCCGTAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAATCACCATCAGCGAT 1080
DB 1021 ATTAGCCGTAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAATCACCATCAGCGAT 1080
QY 1081 ATAATAACCCAAACCATTTACTTACAGCATTTGCGCGCTCAAAATCAAGCGGTCAATCTG 1140
DB 1081 ATAATAACCCAAACCATTTACTTACAGCATTTGCGCGCTCAAAATCAAGCGGTCAATCTG 1140
QY 1141 GCGCATATTTTCCCAAGCGGTACATTAATGTCCGTGCTGCCACTATTCGAAACCAA 1200
DB 1141 GCGCATATTTTCCCAAGCGGTACATTAATGTCCGTGCTGCCACTATTCGAAACCAA 1200
QY 1201 GGTAAACTTTCTGCTGATTTGTAAGCAAGATATAAGCGGCAATATTTCTTTCCGCC 1260
DB 1201 GGTAAACTTTCTGCTGATTTGTAAGCAAGATATAAGCGGCAATATTTCTTTCCGCC 1260
QY 1261 AAGAGGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
DB 1261 AAGAGGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
QY 1321 GCGAAGCTGATGATTACAGCGGATAAGTCAATTAATAAACAGGTGCGAGTTATCGACCTT 1380
DB 1321 GCGAAGCTGATGATTACAGCGGATAAGTCAATTAATAAACAGGTGCGAGTTATCGACCTT 1380
QY 1381 TCAGTAAAGAGGGGAGAACTTTACCTTGGCGGTGACGAGCGGGGGAAGGTAAANAAC 1440
DB 1381 TCAGTAAAGAGGGGAGAACTTTACCTTGGCGGTGACGAGCGGGGGAAGGTAAANAAC 1440
QY 1441 GGCATTCAATTAGCAAGAAACCTCTTTAGAAAAGGCTCAACATCAATGTATCAGGC 1500
DB 1441 GGCATTCAATTAGCAAGAAACCTCTTTAGAAAAGGCTCAACATCAATGTATCAGGC 1500
QY 1501 AAAGAAAAGGGGAGCGGCTATTGTGGGCGATATTGGGTAAATGACGGCAATATT 1560
DB 1501 AAAGAAAAGGGGAGCGGCTATTGTGGGCGATATTGGGTAAATGACGGCAATATT 1560
QY 1561 AACGCTCAAGGTAGTGGTATATCGCTAAAACCGGTGTTTTGTGGAGACATCGGGCAT 1620
DB 1561 AACGCTCAAGGTAGTGGTATATCGCTAAAACCGGTGTTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTTATCCATTGACAGCAATGCAATTTCTTAAACAAAGAGTGGTCTAGACCCCTGAT 1680
DB 1621 TATTTATCCATTGACAGCAATGCAATTTCTTAAACAAAGAGTGGTCTAGACCCCTGAT 1680
QY 1681 GATGTAACAATTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTATAAATGATGAATTC 1740
DB 1681 GATGTAACAATTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTATAAATGATGAATTC 1740
QY 1741 CCAACAGCAGCGGTGAAGCAAGCGACCTTAAAAAATAGCGNACTCAAAACAGCGTA 1800
DB 1741 CCAACAGCAGCGGTGAAGCAAGCGACCTTAAAAAATAGCGNACTCAAAACAGCGTA 1800
QY 1801 ACCAATCAACTATTCAAAATTAATCTGAAAAGCGCTGGACATGAATATAAGCGCATCA 1860
DB 1801 ACCAATCAACTATTCAAAATTAATCTGAAAAGCGCTGGACATGAATATAAGCGCATCA 1860
QY 1861 AGAAAACCTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCCAT 1920
DB 1861 AGAAAACCTACCGTTAATAGCTCAATCAATCAACATCGGAAGCAACTCCCACTTAATTTCCAT 1920
QY 1921 AGTAAAGGTGAGCGGCGGTTCAGATTGATGGAGATATTACTCTTAAGCGCGGA 1980
DB 1921 AGTAAAGGTGAGCGGCGGTTCAGATTGATGGAGATATTACTCTTAAGCGCGGA 1980
QY 1981 AATTAAACCAATTTATCTGCGGATGGTGTGATTTTATAAAAAATATTACGGTTGATCAG 2040
DB 1981 AATTAAACCAATTTATCTGCGGATGGTGTGATTTTATAAAAAATATTACGGTTGATCAG 2040

QY 2041 GGTTTTTTAAATATTACCGCCGCTTCCGTAGCTTTTGAAGGTGAAATAAACAAACGCGC 2100
DB 2041 GGTTTTTTAAATATTACCGCCGCTTCCGTAGCTTTTGAAGGTGAAATAAACAAACGCGC 2100
QY 2101 GACGCGGCAAAATGCTTAAATTTCTGCCCGACGCGCTGTAAACATTACAGGAGAGGAAAA 2160
DB 2101 GACGCGGCAAAATGCTTAAATTTCTGCCCGACGCGCTGTAAACATTACAGGAGAGGAAAA 2160
QY 2161 GATTTTCAGGGCTTAAACAGGTATCTTTAAACGGAAGCGGTAAAGGTCTGAATATCATTTCA 2220
DB 2161 GATTTTCAGGGCTTAAACAGGTATCTTTAAACGGAAGCGGTAAAGGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATATTTAAACCCACAATCTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280
DB 2221 TCAGTGAATATTTAAACCCACAATCTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACGCGCATGATTCGCACCTGG 2340
DB 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACGCGCATGATTCGCACCTGG 2340
QY 2341 AACGTCAGTCTCTTAATCTAGAGACAGGCGCAAAATTTTACCTTTATTAATACATTTCA 2400
DB 2341 AACGTCAGTCTCTTAATCTAGAGACAGGCGCAAAATTTTACCTTTATTAATACATTTCA 2400
QY 2401 AGCAATAGCAAAAGCTTAAACAAACACAGTATAGAGCTCTGCAGGGGTGAATTTAAGCGC 2460
DB 2401 AGCAATAGCAAAAGCTTAAACAAACACAGTATAGAGCTCTGCAGGGGTGAATTTAAGCGC 2460
QY 2461 GTAATGGCAACATGTCTCAATCTCAAAAGAGGAGGCAAAATTTTACCTTTATTAATACATTTCA 2520
DB 2461 GTAATGGCAACATGTCTCAATCTCAAAAGAGGAGGCAAAATTTTACCTTTATTAATACATTTCA 2520
QY 2521 CCAAAACGAGCAATGACCAACCAACACCTTTACCAATTCGGTTTTAGCCAATATCACA 2580
DB 2521 CCAAAACGAGCAATGACCAACCAACACCTTTACCAATTCGGTTTTAGCCAATATCACA 2580
QY 2581 GCCACTGGTGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGCAGAGGGGCT 2640
DB 2581 GCCACTGGTGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGCAGAGGGGCT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTACGGCGCTAAATTTTACCTTTAAATTTCCCAT 2700
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTACGGCGCTAAATTTTACCTTTAAATTTCCCAT 2700
QY 2701 GTTCGCGGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATAAATGCAACCAATTTCA 2760
DB 2701 GTTCGCGGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATAAATGCAACCAATTTCA 2760
QY 2761 AATTTTCAGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC 2820
DB 2761 AATTTTCAGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC 2820
QY 2821 AATTTTCAGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC 2880
DB 2821 AATTTTCAGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC 2880
QY 2881 AGCAGCAGCATTTACGGGGAATTTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAGCC 2940
DB 2881 AGCAGCAGCATTTACGGGGAATTTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAGCC 2940
QY 2941 AATAAGCCCTTAATCAGCAAAACATAGGGATAGATTATATAAACTTTGGCAGCTTGCTC 3000
DB 2941 AATAAGCCCTTAATCAGCAAAACATAGGGATAGATTATATAAACTTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGAGTTTAAAGTTTAACTGGCAAAATGCAGATATTAAAGGCAATCTCACTATT 3060
DB 3001 GTTAATGGAGTTTAAAGTTTAACTGGCAAAATGCAGATATTAAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACCCGGCAATTTT 3120
DB 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACCCGGCAATTTT 3120
QY 3121 ACCAATAATGGCACTGCCGGAATTTAATAACAAAGAGGTGGTAAAACTTTGGCAATGTT 3180

Db 3121 ACCAATATGGCACTGCGCAATTAATATACACAAAGGAGTGTGTAACCTTGGCAATGTT 3180
QY 3181 ACCAATGATGGTATTATTAACATTTACCACTACCGCTAAAGCGCAACCAAGAGCATCATC 3240
Db 3181 ACCAATGATGGTATTATTAACATTTACCACTACCGCTAAAGCGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAAAGGAAGCTTAAATATTACACAGTAGTAAATATGATGCT 3300
Db 3241 GCGGAGATATAATCAACAAAAAGGAAGCTTAAATATTACACAGTAGTAAATATGATGCT 3300
QY 3301 GAATCCAAATTCGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTCTTCCGAT 3360
Db 3301 GAATCCAAATTCGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTCTTCCGAT 3360
QY 3361 AAAATTAATTCACCAACAGATAACAAATCAAAAAGGCTATTGATGGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATTCACCAACAGATAACAAATCAAAAAGGCTATTGATGGAGAGGACTCTAGT 3420
QY 3421 TCAGATCGCAAGTAATGCCAAGCTAACTATTAAAAACCAAGAAATTTGAAATTCACAGAA 3480
Db 3421 TCAGATCGCAAGTAATGCCAAGCTAACTATTAAAAACCAAGAAATTTGAAATTCACAGAA 3480
QY 3481 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
Db 3481 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
QY 3541 ACTATTGCAACAGTAATGACGGTTAACAGCGGTGCCGAAGCCAAACAGTAACCTTTTAAAC 3600
Db 3541 ACTATTGCAACAGTAATGACGGTTAACAGCGGTGCCGAAGCCAAACAGTAACCTTTTAAAC 3600
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAACATGTGACACTTAAATAGCAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAACATGTGACACTTAAATAGCAAGTG 3660
QY 3661 AAAACATCTAGCAAGTGGCGAGCTGAAGCAATAGCGCAACAGTACCGGCTTAACT 3720
Db 3661 AAAACATCTAGCAAGTGGCGAGCTGAAGCAATAGCGCAACAGTACCGGCTTAACT 3720
QY 3721 ATTACTGCAAAAATGTAGAGTAAACAAAGATATTACTTCTCAAAACAGTAAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAGTAAACAAAGATATTACTTCTCAAAACAGTAAATATC 3780
QY 3781 ACCCGTCGGAAGGTTTACCACCAACAGCGGTGACCAATTAACGCAACAAATGGCAAA 3840
Db 3781 ACCCGTCGGAAGGTTTACCACCAACAGCGGTGACCAATTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACAAACCAACAGGTGATATACGCGGTACGATTTCGCGTAAACACGCTAAGT 3900
Db 3841 GCAAGTATTACAAACCAACAGGTGATATACGCGGTACGATTTCGCGTAAACACGCTAAGT 3900
QY 3901 GTTACGCGACTGGTATTAACTAACTAAATTCGCGCTCAAAAATTTGAAGCGAAATCGGGT 3960
Db 3901 GTTACGCGACTGGTATTAACTAACTAAATTCGCGCTCAAAAATTTGAAGCGAAATCGGGT 3960
QY 3961 GAGGCTAATGTAAAGTGCACAGTACAAATTTGCGGTGACAAATTTCCGGTAAATAGGTA 4020
Db 3961 GAGGCTAATGTAAAGTGCACAGTACAAATTTGCGGTGACAAATTTCCGGTAAATAGGTA 4020
QY 4021 AATGTTACGGCAACCGTGGCGATTATACAGTTGGGAATGGCGGAGAAATTAATCCGACA 4080
Db 4021 AATGTTACGGCAACCGTGGCGATTATACAGTTGGGAATGGCGGAGAAATTAATCCGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAAACCGCAACAGGGAATACCTTGACTACTGAAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAAACCGCAACAGGGAATACCTTGACTACTGAAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTTGGCTCAGAAATGGTAGCATCGCAGGAAGC 4200
Db 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTTGGCTCAGAAATGGTAGCATCGCAGGAAGC 4200
QY 4201 ATTAATGCTGCTAAATGTGACATTAATATCTACAGGCACCTTAACCAACCGTGGCAGGCTCG 4260

Db 4201 ATTAATGCTGCTAAATGTGACATTAATATCTACAGGCACCTTTAAACCAACCGTGGCAGGCTCG 4260
QY 4261 GATATTAAGCAACAGCGGCACCTTTGGTATTATTAACGCAAAAGATGCTTAAGCTAANTGTT 4320
Db 4261 GATATTAAGCAACAGCGGCACCTTTGGTATTATTAACGCAAAAGATGCTTAAGCTAANTGTT 4320
QY 4321 GATGATCAGGTGATGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATGATCAGGTGATGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCTACTGGGGATTTAAACACACAGTAAATGGGTTA 4440
Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCTACTGGGGATTTAAACACACAGTAAATGGGTTA 4440
QY 4441 AATATCATTTCCAAAGATGTTAGAAACACTGTGCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500
Db 4441 AATATCATTTCCAAAGATGTTAGAAACACTGTGCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500
QY 4501 AATATATCCAGCGAGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACCGCTTCTT 4560
Db 4501 AATATATCCAGCGAGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACCGCTTCTT 4560
QY 4561 GAAAAAGTAAAAGATTATCTGATGAAGAAAGAGAAACATTTAGCTTAAGT 4620
Db 4561 GAAAAAGTAAAAGATTATCTGATGAAGAAAGAGAAACATTTAGCTTAAGT 4620
QY 4621 GCTGTACGTTTTTGTGAGCCAAATAATACTACAGTCAATACAGCAAAATGAATTTACA 4680
Db 4621 GCTGTACGTTTTTGTGAGCCAAATAATACTACAGTCAATACAGCAAAATGAATTTACA 4680
QY 4681 ACCAGACCGCTCAAGTCAAGTGAATTTCTGAAGSTAAAGCGGTGTTCTCAAGTGGTAAT 4740
Db 4681 ACCAGACCGCTCAAGTCAAGTGAATTTCTGAAGSTAAAGCGGTGTTCTCAAGTGGTAAT 4740
QY 4741 GCGCGACAGGATGTACCAATGTGCTGACGATGGACGCGGTAGTCAAGTAAATTTGACAAG 4800
Db 4741 GCGCGACAGGATGTACCAATGTGCTGACGATGGACGCGGTAGTCAAGTAAATTTGACAAG 4800
QY 4801 GTAGATTTCATCTCGCAATGAAGTCAATTTATTTTCTGATTTATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTCATCTCGCAATGAAGTCAATTTATTTTCTGATTTATTTACTGTGGGTTAAA 4860
QY 4861 GTTCAGTACGGGCTTTACCCATCTTGAAGAAATTTAGGAGAAATCAATAAAGTATTTT 4920
Db 4861 GTTCAGTACGGGCTTTACCCATCTTGAAGAAATTTAGGAGAAATCAATAAAGTATTTT 4920
QY 4921 AACAGGTTATTATTATG 4937
Db 4921 AACAGGTTATTATTATG 4937

RESULT 8
AR032350
LOCUS
DEFINITION
AR032350
ACCESSION
AR032350.1
VERSION
AR032350.1
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 9323)
Barenkamp, S. J. and St. Geme, J. William, III.
High molecular weight surface proteins of non-typeable haemophilus
Patent: US 5869065-A 6 09-FEB-1999;
JOURNAL
FEATURES
source
location/Qualifiers
i
BASE COUNT 3134 a 1829 c 1832 g 2528 t
ORIGIN
Query Match 97.6%; Score 4820.2; DB 6; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;

Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATACAGATAATAAATAAATCAAGATTTTGTGATGACACAAACAATAATACAA 60
Db 441 TAAATACAGATAATAAATAAATCAAGATTTTGTGATGACACAAACAATAATACAA 500
QY 61 CACCTTTTGTGAGTCTATATGCAATATTTTAAATAAATAGTATAAATCCGCCATATA 120
Db 501 CACCTTTTGTGAGTCTATATGCAATATTTTAAATAAATAGTATAAATCCGCCATATA 560
QY 121 AATGGTATATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 180
Db 561 AATGGTATATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 620
QY 181 CTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTT 240
Db 621 CTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTT 680
QY 241 CACATGAATGATGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 681 CACATGAATGATGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
QY 301 GAACGCAATGATAAAGTAATTTAATTTCACTTAACCTTAGGAGAGAAATATGAACAAG 360
Db 741 GAACGCAATGATAAAGTAATTTAATTTCACTTAACCTTAGGAGAGAAATATGAACAAG 800
QY 361 ATATATCTCTCAAAATTCAGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
Db 801 ATATATCTCTCAAAATTCAGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 860
QY 421 CGGGTGTGACCATTCACAGAGAAAGGCTTCCGCTATGTACTATCTTTCATCTTTCATCTT 480
Db 861 CGGGTGTGACCATTCACAGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 920
QY 481 CACTTAGCGTTAAAGCCACTTTCGCTATGTACTATCTTTCATCTTTCATCTTTCATCTT 540
Db 921 CACTTAGCGTTAAAGCCACTTTCGCTATGTACTATCTTTCATCTTTCATCTTTCATCTT 980
QY 541 CAATCTGTTTACGAGCGGCTTACAGGAATGGATGTAGTACAGGACAGGACAGGACAGT 600
Db 981 CAATCTGTTTACGAGCGGCAATTTAATCTGACCAAAATGAAATGGTGAGTGTATTTAC 1040
QY 601 CAAGTAGATGTAATAAACCATTATCGCAACAGTGTGACGCTATCATTTAATTTGAAA 660
Db 1041 AAGAAACAAGTAATAAACCATTATCGCAACAGTGTGACGCTATCATTTAATTTGAAA 1100
QY 661 CAATTTACATCGACCAAAATGAAATGGTGAGTGTATTTACAGAAACAACAACCTCGGCC 720
Db 1101 CAATTTACATCGACCAAAATGAAATGGTGAGTGTATTTACAGAAACAACAACCTCGGCC 1160
QY 721 GTATTTCAACCGTGTATACATCTTAACCAATCTCCCAATTTAAAGGATTTTATGATTTAAC 780
Db 1161 GTATTTCAACCGTGTATACATCTTAACCAATCTCCCAATTTAAAGGATTTTATGATTTAAC 1220
QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATACATAGTATAAGAGCAATTTAATTAAC 840
Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGGTATACATAGTATAAGAGCAATTTAATTAAC 1280
QY 841 ACTAATGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGGCGCGTAAT 900
Db 1281 ACTAATGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGGCGCGTAAT 1340
QY 901 TTCACCTTCGAGCAAAACCAAGATAAAGGCTCGCTGCAATTTGTGAATCAGGTTTAATTT 960
Db 1341 TTCACCTTCGAGCAAAACCAAGATAAAGGCTCGCTGCAATTTGTGAATCAGGTTTAATTT 1400
QY 961 ACTGTGCGTAAAGACGCGATGTAAATCTTTATTTGGTGGCAAGTGAATAAAGAGGGTGTG 1020
Db 1401 ACTGTGCGTAAAGACGCGATGTAAATCTTTATTTGGTGGCAAGTGAATAAAGAGGGTGTG 1460
QY 1021 ATTAGCGTAATGGTGCAGCATTTCTTTTACTCGAGGGCAAAATACCATTCAGCGAT 1080
Db 1461 ATTAGCGTAATGGTGCAGCATTTCTTTTACTCGAGGGCAAAATACCATTCAGCGAT 1520

QY 1081 ATATATAACCCACCATTACTTACAGCATTTGCCGGCCTGAAATAGACGGTCAATCTG 1140
Db 1521 ATATATAACCCACCATTACTTACAGCATTTGCCGGCCTGAAATAGACGGTCAATCTG 1580
QY 1141 GCGGATATTTTTCGCAAAAGCGGTAACTTAATGTCGCTGCCACTATTTCCAAACCAA 1200
Db 1581 GCGGATATTTTTCGCAAAAGCGGTAACTTAATGTCGCTGCCACTATTTCCAAACCAA 1640
QY 1201 GGTAAACTTTCGCTGATTTCTGTAGCAAGATTAAGGCGCAATATTTCTTTTCGCGC 1360
Db 1641 GGTAAACTTTCGCTGATTTCTGTAGCAAGATTAAGGCGCAATATTTCTTTTCGCGC 1700
QY 1261 AAAGAGGGTGAAGCGGAAATTTGCGGTGTAAATTTCCGCTCAAAATCAGCAAGCTAAGGC 1320
Db 1701 AAAGAGGGTGAAGCGGAAATTTGCGGTGTAAATTTCCGCTCAAAATCAGCAAGCTAAGGC 1760
QY 1321 GGCAAGCTGATGATTACAGCGGATTAAGTCAATTAACACAGGTGACGTTATTCGACCTT 1380
Db 1761 GGCAAGCTGATGATTAAAGTCCGATTAAGTCAATTAACACAGGTGACGTTATTCGACCTT 1820
QY 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTGGGTGACGAGCGGCGGAGGATAAAGC 1440
Db 1821 TCAGGTAAAGAGGGGAGAACTTTACCTTGGGTGACGAGCGGCGGAGGATAAAGC 1880
QY 1441 GGCATTCAATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATCTATCAGGC 1500
Db 1881 GGCATTCAATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATCTATCAGGC 1940
QY 1501 AAAGAAAAGGGCGGCGCTATTGTGGGCGATATTGCGTTAATTTGACGGCAATATT 1560
Db 1941 AAAGAAAAGGGCGGCGCTATTGTGGGCGATATTGCGTTAATTTGACGGCAATATT 2000
QY 1561 AACGCTCAAGGTAGTGGTATCGCTAAACCGGTGGTTTTGTGGAGACATCGGGGAT 1620
Db 2001 AACGCTCAAGGTAGTGGTATCGCTAAACCGGTGGTTTTGTGGAGACATCGGGGAT 2060
QY 1621 TATTATCCATTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTGTAGACCTGAT 1680
Db 2061 TATTATCCATTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTGTAGACCTGAT 2120
QY 1681 GATGTAAACATTTGAAGCGGAGAGCCCTTCGCAATATACCGGTATTAATGATGAATTC 1740
Db 2121 GATGTAAACATTTGAAGCGGAGAGCCCTTCGCAATATACCGGTATTAATGATGAATTC 2180
QY 1741 CCAACAGCAGCGGTGAAGCAAGCAAGCCCTTAAACAAAAGAGTGGTTGTAGACCTGAT 1800
Db 2181 CCAACAGCAGCGGTGAAGCAAGCAAGCCCTTAAACAAAAGAGTGGTTGTAGACCTGAT 2240
QY 1801 ACCAATACACTATTTTCAATTTCTGAAAACGCTGGACAATGAATATAACGGCATCA 1860
Db 2241 ACCAATACACTATTTTCAATTTCTGAAAACGCTGGACAATGAATATAACGGCATCA 2300
QY 1861 AGAAACTTACCGTTTAAATAGCTCAATCAACATCGGAACCACTCCCACTTAATTTCTCCAT 1920
Db 2301 AGAAACTTACCGTTTAAATAGCTCAATCAACATCGGAACCACTCCCACTTAATTTCTCCAT 2360
QY 1921 AGTAAAGTCAAGCTGGCGGAGGCTTCAGATTGATGAGATATTACTTCTTAAAGCGGGA 1980
Db 2361 AGTAAAGTCAAGCTGGCGGAGGCTTCAGATTGATGAGATATTACTTCTTAAAGCGGGA 2420
QY 1981 AATTTAACCATTATTCTGGCGGATGGTTCATATAAATAATTTACGCTTGATCAG 2040
Db 2421 AATTTAACCATTATTCTGGCGGATGGTTCATATAAATAATTTACGCTTGATCAG 2480
QY 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGTGGAAATAACAAAGCAGGC 2100
Db 2481 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGTGGAAATAACAAAGCAGGC 2540
QY 2101 GACCGGCAATGCTAAAATTTGTCGCCAGGCGACTGTAAACATTACAGGAGAGGAAAA 2160
Db 2541 GACCGGCAATGCTAAAATTTGTCGCCAGGCGACTGTAAACATTACAGGAGAGGAAAA 2600

QY 2161 GATTCAGGGCTAACACCTATCTTTAAAGGAGGCTAAAGGCTGAATATCATTTCA 2220
 Db 2601 GATTCAGGGCTAACACCTATCTTTAAAGGAGGCTAAAGGCTGAATATCATTTCA 2660
 QY 2221 TCAGTGAATTAATTAACCCACATCTTAGTGCCACAAATTAACATATCTGGGAATATAACA 2280
 Db 2661 TCAGTGAATTAATTAACCCACATCTTAGTGCCACAAATTAACATATCTGGGAATATAACA 2720
 QY 2281 ATTAACCAACTACGAAAGAACCTCGTATTTGGCAACACAGCCATGATTCGCACTGG 2340
 Db 2721 ATTAACCAACTACGAAAGAACCTCGTATTTGGCAACACAGCCATGATTCGCACTGG 2780
 QY 2341 AACGTCAGTCTCTTAATCTAGAGACAGCGCAAAATTTTAAATACATTTCA 2400
 Db 2781 AACGTCAGTCTCTTAATCTAGAGACAGCGCAAAATTTTAAATACATTTCA 2840
 QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAAGCTCTGCAGGGTGAATTTAAAGGC 2460
 Db 2841 AGCAATAGCAAGGCTTAACACACAGTATAGAAGCTCTGCAGGGTGAATTTAAAGGC 2900
 QY 2461 GTAAATGCAACATCTCATCTCAAGAGAGGAGGAAAGTTAAATTCAAATTAATA 2520
 Db 2901 GTAAATGCAACATCTCATCTCAAGAGAGGAGGAAAGTTAAATTCAAATTAATA 2960
 QY 2521 CCNAACGAGAACATGACACAGCAACCTTTACCAATTCGGTTTTAGCCAAATATACA 2580
 Db 2961 CCNAACGAGAACATGACACAGCAACCTTTACCAATTCGGTTTTAGCCAAATATACA 3020
 QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGATATATATGCCAACCATTTCTGGCAGAGGGCT 2640
 Db 3021 GCCACTGGTGGGGCTCTGTTTTTTTGATATATATGCCAACCATTTCTGGCAGAGGGCT 3080
 QY 2641 GAGTAAAAATGAGTCAAAATTAATATCTCTAACGGCGCTAAATTTTACCTTTAAATTCGCCAT 2700
 Db 3081 GAGTAAAAATGAGTCAAAATTAATATCTCTAACGGCGCTAAATTTTACCTTTAAATTCGCCAT 3140
 QY 2701 GTTCGGCGGATGACGCTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATTC 2760
 Db 3141 GTTCGGCGGATGACGCTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATTC 3200
 QY 2761 AATTCAGGCTCAGACAGAGAAAGATTTTTATGACGGGTACGACGCAATGCGATC 2820
 Db 3201 AATTCAGGCTCAGACAGAGAAAGATTTTTATGACGGGTACGACGCAATGCGATC 3260
 QY 2821 AATTCAGGCTCAGACAGAGAAAGATTTTTATGACGGGTACGACGCAATGCGATC 2880
 Db 3261 AATTCAGGCTCAGACAGAGAAAGATTTTTATGACGGGTACGACGCAATGCGATC 3320
 QY 2881 AGCAGCAGCATTAACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGGCTAGAAGCC 2940
 Db 3321 AGCAGCAGCATTAACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGGCTAGAAGCC 3380
 QY 2941 AATTAACGGCCCTAATCAGCAAAACATTAAGGATAGAGTTATAAACTTGGCAGCTTGCTC 3000
 Db 3381 AATTAACGGCCCTAATCAGCAAAACATTAAGGATAGAGTTATAAACTTGGCAGCTTGCTC 3440
 QY 3001 GTTAATGGGATTTAAGTTAACTGGCGAAATGCGATATTAAGGCAATCTCACTATT 3060
 Db 3441 GTTAATGGGATTTAAGTTAACTGGCGAAATGCGATATTAAGGCAATCTCACTATT 3500
 QY 3061 TCAGAAGCGCCACTTTTAAAGAAAGACTAGAGATACCCCTAAATATACCGGCAATTTT 3120
 Db 3501 TCAGAAGCGCCACTTTTAAAGAAAGACTAGAGATACCCCTAAATATACCGGCAATTTT 3560
 QY 3121 ACCAATATGGCACTCCGCAATTAATATAACACAGAGAGTGGTAAACCTTGGCAATGTT 3180
 Db 3561 ACCAATATGGCACTCCGCAATTAATATAACACAGAGAGTGGTAAACCTTGGCAATGTT 3620
 QY 3181 ACCAATGATGGTATTTAAACATTTACCACTACGCTAAACGCAACCAAGCAATCATC 3240
 Db 3621 ACCAATGATGGTATTTAAACATTTACCACTACGCTAAACGCAACCAAGCAATCATC 3680
 QY 3241 GCGGGAGATATATACCAAAAAAGGAGCTTAAATATTACACAGCATTAATGATGCT 3300

Db 3681 GCGGGAGATATATACCAAAAAAGGAGCTTAAATATTACACAGTAAATGATGCT 3740
 QY 3301 GAAATCCAAATTCGGGCAATATCTCGCAAAAAGAGGCAACCTCAGGATTTCTTCGGAT 3360
 Db 3741 GAAATCCAAATTCGGGCAATATCTCGCAAAAAGAGGCAACCTCAGGATTTCTTCGGAT 3800
 QY 3361 AAAATTAATATCAACCAACAGATAACAATCAAAAAGGATTTGATGGAGAGGACTCTAGT 3420
 Db 3801 AAAATTAATATCAACCAACAGATAACAATCAAAAAGGATTTGATGGAGAGGACTCTAGT 3860
 QY 3421 TCAGATCCGCAACAGTAAATGCAACCTTAATTAATAACCAAGAAATTTGAAATTCACAGAA 3480
 Db 3861 TCAGATCCGCAACAGTAAATGCAACCTTAATTAATAACCAAGAAATTTGAAATTCACAGAA 3920
 QY 3481 GACCTAAGTATTTTCAAGTTTCAATTAAGCAGAGATTTACGCCAAAGATGTTAGAGATTTA 3540
 Db 3921 GACCTAAGTATTTTCAAGTTTCAATTAAGCAGAGATTTACGCCAAAGATGTTAGAGATTTA 3980
 QY 3541 ACTATTTGGCAACAGTAAATGACGGTAAACAGCGGTCGCCAAAGCAACAGTAACCTTTAAC 3600
 Db 3981 ACTATTTGGCAACAGTAAATGACGGTAAACAGCGGTCGCCAAAGCAACAGTAACCTTTAAC 4040
 QY 3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCACAATGTGACACTTAAATFAGCAAGTG 3660
 Db 4041 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCACAATGTGACACTTAAATFAGCAAGTG 4100
 QY 3661 AAAACATCTAGCAGCAATGCGGACGTGAAAGCAATAGCGACACAGATACCGGCTTAAC 3720
 Db 4101 AAAACATCTAGCAGCAATGCGGACGTGAAAGCAATAGCGACACAGATACCGGCTTAAC 4160
 QY 3721 ATTACTGCAAAAATGTAGAAATTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780
 Db 4161 ATTACTGCAAAAATGTAGAAATTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 4220
 QY 3781 ACCGCTCGCAAAAGGTTACCCACAGCAGGCTCGACCATTAACGCAACAAATGCGCAA 3840
 Db 4221 ACCGCTCGCAAAAGGTTACCCACAGCAGGCTCGACCATTAACGCAACAAATGCGCAA 4280
 QY 3841 GCAAGTATTAACCAAAACAGGTGATACAGCGGTACGATTTCCGGTACACCGGTAGT 3900
 Db 4281 GCAAGTATTAACCAAAACAGGTGATACAGCGGTACGATTTCCGGTACACCGGTAGT 4340
 QY 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATTCGGGCTCAAAAATTTGAAGGCAATCGG 3960
 Db 4341 GTTAGCGGACTGGTGAATTTAACCACTAAATTCGGGCTCAAAAATTTGAAGGCAATCGG 4400
 QY 3961 GAGGCTAATGTAACAGTGCAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATACGTA 4020
 Db 4401 GAGGCTAATGTAACAGTGCAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATACGTA 4460
 QY 4021 AATGTTACGCAACGCTGGCGATTTTAAACAGTTGGGAATGGCGCAGAAATTTAATGCGACA 4080
 Db 4461 AATGTTACGCAACGCTGGCGATTTTAAACAGTTGGGAATGGCGCAGAAATTTAATGCGACA 4520
 QY 4081 GAAGGAGCTCAACCTTTAACGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
 Db 4521 GAAGGAGCTCAACCTTTAACGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4580
 QY 4141 ATCACTCACTAAGGCTAGGTAGACCTCTTGCTCAGAAATGTTAGTATCGCAGGAGC 4200
 Db 4581 ATCACTCACTAAGGCTAGGTAGACCTCTTGCTCAGAAATGTTAGTATCGCAGGAGC 4640
 QY 4201 ATTAATGCTGCTAATGTGACATTAATTAACAGCAACCTTTAACCCCGTGGCAGGCTCG 4260
 Db 4641 ATTAATGCTGCTAATGTGACATTAATTAACAGCAACCTTTAACCCCGTGGCAGGCTCG 4700
 QY 4261 GATATTAAGCAACACGCGCACCTTGCTTATTAAACGCAAAAGATGCTAAGCTAAATGCT 4320
 Db 4701 GATATTAAGCAACACGCGCACCTTGCTTATTAAACGCAAAAGATGCTAAGCTAAATGCT 4760
 QY 4321 GATGCTACAGTATAGTACAGAGTAAATGCAAGTAAATGCAAGCGCTCTGCTAGTGTG 4380

Db 4761 GATGATCAGGTGATAGTACAGAAAGTGAATGCAGTCAACGACTGGGATTTGGTAGTGTG 4820
QY 4381 ACTGGCGCAACCTCAAGCAGTGTGAATATCACTACCTGGGATTTAAACACAGTAATGGGTTA 4440
Db 4821 ACTGGCGCAACCTCAAGCAGTGTGAATATCACTACCTGGGATTTAAACACAGTAATGGGTTA 4880
QY 4441 AATATCATTTTCAAGAGTGGTAGAACACACTGTCGCGTTTAAGAGGCAAGAAATGAGGTG 4500
Db 4881 AATATCATTTTCAAGAGTGGTAGAACACACTGTCGCGTTTAAGAGGCAAGAAATGAGGTG 4940
QY 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTGAATTAAGAGGCAAGAAATGAGGTG 4560
Db 4941 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTGAATTAAGAGGCAAGAAATGAGGTG 5000
QY 4561 GAAAAGTAAAGATTTATCTGATGAAGAGAGAGAAACATTAGCTTAACTTTGGTGTAAAGT 4620
Db 5001 GAAAAGTAAAGATTTATCTGATGAAGAGAGAGAAACATTAGCTTAACTTTGGTGTAAAGT 5060
QY 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTTAGAGTCAATACAGTCAATACAAATTAAGT 4680
Db 5061 GCTGTACGTTTGTGAGCCAAATATACAAATTTAGAGTCAATACAGTCAATACAAATTAAGT 5120
QY 4681 ACCGACCGCTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAT 4740
Db 5121 ACCGACCGCTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAT 5180
QY 4741 GCGCACGAGTATGACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACAAAG 4800
Db 5181 GCGCACGAGTATGACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACAAAG 5240
QY 4801 GTAGATTTCACTCCGCAATGAAGTCAATTTATTTTCGATATTATTTACGTTGGTGTAAA 4860
Db 5241 GTAGATTTCACTCCGCAATGAAGTCAATTTATTTTCGATATTATTTACGTTGGTGTAAA 5300
QY 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAAGAAATTTACGAGAAATACAAATAAAGTATTTT 4920
Db 5301 GTTCAGTACGGGCTTTACCCATCTTTGTAAGAAATTTACGAGAAATACAAATAAAGTATTTT 5360
QY 4921 AACAGGTTATTATTATG 4937
Db 5361 AACAGGTTATTATTATG 5377

RESULT 9
AR083957
LOCUS AR083957 9323 bp DNA linear PAT 01-SEP-2000
DEFINITION Sequence 6 from patent US 5977336.
ACCESSION AR083957
VERSION AR083957.1 GI:10010728
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 9323)
AUTHORS Barenkamp, S.J.
TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 5977336-A 6 02-NOV-1999;
FEATURES
source
1. .9323
/organism="unknown"
BASE COUNT 3134 a 1829 c 1832 g 2528 t
ORIGIN

Query Match 97.6%; Score 4820.2; DB 6; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGATAATAAATAAATCAAGATTTTGTGATGACAAACAAATTAACA 60
Db 441 TAAATATACAGATAATAAATAAATCAAGATTTTGTGATGACAAACAAATTAACA 500
QY 61 CACCTTTTTCAGCTGTATATGCAATATTTTAAAAAATAGTAAATCCGCCATATAA 120
|||||

Db 501 CACCTTTTTCAGCTGTATATGCAAAATATTTAAAAAATAGTAAATCCGCCATATAA 560
QY 121 AATGATATAATCTTTCATCTTTTCATCTTTAACTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
|||||
Db 561 AATGATATAATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
|||||
Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680
QY 241 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 681 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
QY 301 GAACGCAATATGAATGAATTAATTTTCAACTTAACCTTAGGAGAGAAATATGAACAAG 360
Db 741 GAACGCAATATGAATGAATTAATTTTCAACTTAACCTTAGGAGAGAAATATGAACAAG 800
QY 361 ATATATCGTCTCAAAATTCAGCAAAACGCCGTGAATGCTTTGCTGTCTGTCTGAATTTGGCA 420
Db 801 ATATATCGTCTCAAAATTCAGCAAAACGCCGTGAATGCTTTGCTGTCTGTCTGAATTTGGCA 860
QY 421 CGGGGTTGTGACCATTCACACAGAAAGGCTTCCGCTATGTTACTATCTTTAGGTGTAAAC 480
Db 861 CGGGGTTGTGACCATTCACACAGAAAGGCGAGCAGCAAAAACCTGCTCGCATGAAGTGCCT 920
QY 481 CACTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAAACATCTATTCCA 540
Db 921 CACTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAAACATCTATTCCA 980
QY 541 CAATCTGTTTTAGCAAGCGGCTTACAAGAAATGGATGTAGTACACGGGACAGCCACTATG 600
Db 981 CAATCTGTTTTAGCAAGCGGCAATTTAACTCGACCAAAATGAAATGCTGCGATGTTTAC 1040
QY 601 CAAGTAGATGTTAAATAAAGCAATTCAGCAACAGTGTGACGCTATCAATTAATTTGGAAA 660
Db 1041 AAGAAAACAGTAATAAAGCAATTCAGCAACAGTGTGACGCTATCAATTAATTTGGAAA 1100
QY 661 CAATTTAACTCGACCAAAATGAAATGCTGCAAGTGTGTTTACAGAAAACAAACAACTCCGCC 720
Db 1101 CAATTTAACTCGACCAAAATGAAATGCTGCAAGTGTGTTTACAGAAAACAAACAACTCCGCC 1160
QY 721 GTATTCACCGTGTACATCTAACCAATCTCCCAATTTAAAGGAGTATTTAGATTTCTAAC 780
Db 1161 GTATTCACCGTGTACATCTAACCAATCTCCCAATTTAAAGGAGTATTTAGATTTCTAAC 1220
QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATCACAAATAGGTAAAGACGCAATTAATTAAC 840
Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGGTATCACAAATAGGTAAAGACGCAATTAATTAAC 1280
QY 841 ACTAATGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAAT 900
Db 1281 ACTAATGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAAT 1340
QY 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGTGTGCTGAAATTTGTAATCAGCGTTTAAAT 960
Db 1341 TTCACCTTCGAGCAAAACCAAGATTAAGCGTGTGCTGAAATTTGTAATCAGCGTTTAAAT 1400
QY 961 ACTGTCGGTAAAGAGCGGAGTGAATCTTATTTGTCGCAAGTGAAGAACAGGAGGTGTG 1020
Db 1401 ACTGTCGGTAAAGAGCGGAGTGAATCTTATTTGTCGCAAGTGAAGAACAGGAGGTGTG 1460
QY 1021 ATTACGCTAAATGTTGGCAGCATTTCTTTTACTCGCAGGCAAAAATCACCATACCGAT 1080
Db 1461 ATTACGCTAAATGTTGGCAGCATTTCTTTTACTCGCAGGCAAAAATCACCATACCGAT 1520
QY 1081 ATAATAACCAACCATTTACTTACAGCATTCGCGGCTGAAATTAAGACGGGTCAATCTG 1140
Db 1521 ATAATAACCAACCATTTACTTACAGCATTCGCGGCTGAAATTAAGACGGGTCAATCTG 1580
QY 1141 GGCATATTTTGGCAAGGCGGTAAACATTAATGTCCTGCTGCCCATATTTCGAAACCAA 1200
Db 1581 GGCATATTTTGGCAAGGCGGTAAACATTAATGTCCTGCTGCCCATATTTCGAAACCAA 1640
|||||

us-10-092-880-3.ige

Tue Mar 25 08:39:00 2003

Qy	1201	GGTAACATTTCTGCTGATTCCTGTAAGCAAGATAAAAGCGGCAATATTTGTTCTTCCGCC	1260
Db	1641	GGTAACATTTCTGCTGATTCCTGTAAGCAAGATAAAAGCGGCAATATTTGTTCTTCCGCC	1700
Qy	1261	AAAGAGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC	1320
Db	1701	AAAGAGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC	1760
Qy	1321	GGCAAGCTGATGATTTACAGCGGATTAAGTCAATTTAAACAGAGTGGCATTTATCGACCTT	1380
Db	1761	GGCAAGCTGATGATTAAGTCCGATTAAGTCAATTTAAACAGAGTGGCATTTATCGACCTT	1820
Qy	1381	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGTAAAC	1440
Db	1821	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGTAAAC	1880
Qy	1441	GGCAATTTAGCAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC	1500
Db	1881	GGCAATTTAGCAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC	1940
Qy	1501	AAAGAAAAAGCGGACGCGCTATTGTGTGGGCGGATATTGCGTTAATTGACGGCAATAT	1560
Db	1941	AAAGAAAAAGCGGACGCGCTATTGTGTGGGCGGATATTGCGTTAATTGACGGCAATAT	2000
Qy	1561	AACGCTCAAGGTAGTGGTGATATCGCTTAAACACCGGTGGTTTTGTGGAGACATCGGGGCAT	1620
Db	2001	AACGCTCAAGGTAGTGGTGATATCGCTTAAACACCGGTGGTTTTGTGGAGACATCGGGGCAT	2060
Qy	1621	TATTTATCCATTGACGCAATGCAATTTGTTAAAAACAAAGAGTGGTTGCTAGACCCCTGAT	1680
Db	2061	TATTTATCCATTGACGCAATGCAATTTGTTAAAAACAAAGAGTGGTTGCTAGACCCCTGAT	2120
Qy	1681	GATGTAAACAATTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTATATAATGATGAATTC	1740
Db	2121	GATGTAAACAATTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTATATAATGATGAATTC	2180
Qy	1741	CCACAGCGCGGTGAAGCAAGGACCCCTAAAAAATAATACGAACTCAAAACACGCTA	1800
Db	2181	CCACAGCGCGGTGAAGCAAGGACCCCTAAAAAATAATACGAACTCAAAACACGCTA	2240
Qy	1801	ACCAATACAACTATTTCAATTTCTGAAGACGCTTGAACCAATCAATATTAACGGCATCA	1860
Db	2241	ACCAATACAACTATTTCAATTTCTGAAGACGCTTGAACCAATCAATATTAACGGCATCA	2300
Qy	1861	AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATCTCCAT	1920
Db	2301	AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATCTCCAT	2360
Qy	1921	AGTAAGGTACGCGTGGCGGAGCGCTCAGATTGATGGAGATTAATTTACGTTCAATCAG	1980
Db	2361	AGTAAGGTACGCGTGGCGGAGCGCTCAGATTGATGGAGATTAATTTACGTTCAATCAG	2420
Qy	1981	AATTTAACCACTTATTTCTGGCGGATGGTTGATTTCAATATAAATAATTTACGTTCAATCAG	2040
Db	2421	AATTTAACCACTTATTTCTGGCGGATGGTTGATTTCAATATAAATAATTTACGTTCAATCAG	2480
Qy	2041	GGTTTTTAAATTAATACCGCGCTCCGTTAGCTTTTGAAGGTGGAATTAACAAAGCACGC	2100
Db	2481	GGTTTTTAAATTAATACCGCGCTCCGTTAGCTTTTGAAGGTGGAATTAACAAAGCACGC	2540
Qy	2101	GACGCGCAATGCTAAATTTGCGCCAGGCGCTGTAACCAATTAACAGGAGGGAATA	2160
Db	2541	GACGCGCAATGCTAAATTTGCGCCAGGCGCTGTAACCAATTAACAGGAGGGAATA	2600
Qy	2161	GATTTAGGCTTAACACGATATCTTTAAACGGAACGGTAAAGTCTGAATATCAATTTCA	2220
Db	2601	GATTTAGGCTTAACACGATATCTTTAAACGGAACGGTAAAGTCTGAATATCAATTTCA	2660
Qy	2221	TCAGTGAATTAATTTAACCCCAATCTTTAGTGGCACATTAACATATCTGGGAATAAACA	2280
Db	2661	TCAGTGAATTAATTTAACCCCAATCTTTAGTGGCACATTAACATATCTGGGAATAAACA	2720
Qy	2281	ATTAAACAACTACGAGAAAGAACACCTCTATTGGCAAAACACGACATGATTCGCACTGG	2340
Db	2721	ATTAAACAACTACGAGAAAGAACACCTCTATTGGCAAAACACGACATGATTCGCACTGG	2780
Qy	2341	AACGTGAGTCTCTTAATCTAGACAGAGCGCAAAATTTTACCTTTTATTAATACATTTCA	2400
Db	2781	AACGTGAGTCTCTTAATCTAGACAGAGCGCAAAATTTTACCTTTTATTAATACATTTCA	2840
Qy	2401	AGCAATAGCAAAAGCTTTAAACACACAGTATAGAGCTCTGAGGGTGAATTTTAAACGGC	2460
Db	2841	AGCAATAGCAAAAGCTTTAAACACACAGTATAGAGCTCTGAGGGTGAATTTTAAACGGC	2900
Qy	2461	GTAATAGCAACATGTCTCAATCTCAAGAGAGGAGGAAGTAAATTTCAAAATTTAAA	2520
Db	2901	GTAATAGCAACATGTCTCAATCTCAAGAGAGGAGGAAGTAAATTTCAAAATTTAAA	2960
Qy	2521	CCAAACGAGAACATGAACACAAAGCAACCTTTTACCAATTCGGTTTATAGCCAAATATCA	2580
Db	2961	CCAAACGAGAACATGAACACAAAGCAACCTTTTACCAATTCGGTTTATAGCCAAATATCA	3020
Qy	2581	GCACCTGGTGGGGGCTCTGTTTTTTTGTATATATATGCAACCATCTCTGCGAGAGGGCT	2640
Db	3021	GCACCTGGTGGGGGCTCTGTTTTTTTGTATATATATGCAACCATCTCTGCGAGAGGGCT	3080
Qy	2641	GAGTTAAAAATCAGTGAATTAATATCTCTAACGGCGCTTAATTTTACCTTTAAATTTCCCAT	2700
Db	3081	GAGTTAAAAATCAGTGAATTAATATCTCTAACGGCGCTTAATTTTACCTTTAAATTTCCCAT	3140
Qy	2701	GTTCGCGGCGATGACGCTTTTAAAAATCAACAAAGAGTAAACCATAAATGCAACCAATTTCA	2760
Db	3141	GTTCGCGGCGATGACGCTTTTAAAAATCAACAAAGAGTAAACCATAAATGCAACCAATTTCA	3200
Qy	2761	AATTCAGCTCAGACAGAGAGAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	2820
Db	3201	AATTCAGCTCAGACAGAGAGAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	3260
Qy	2821	AATTCAGCTCAGACAGAGAGAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	2880
Db	3261	AATTCAGCTCAGACAGAGAGAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	3320
Qy	2881	AGCAGCAGATATCGGGGATATTTACTATCGAGAAAGAGCAAAATGTTACGCTAGAAGCC	2940
Db	3321	AGCAGCAGATATCGGGGATATTTACTATCGAGAAAGAGCAAAATGTTACGCTAGAAGCC	3380
Qy	2941	ANTAAACCCCTTAATCAGCAAAACATAGGATAGAGTTTAAACCTTGGCAGCTTGCCTC	3000
Db	3381	ANTAAACCCCTTAATCAGCAAAACATAGGATAGAGTTTAAACCTTGGCAGCTTGCCTC	3440
Qy	3001	GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTAAGGCAATCTCACTATT	3060
Db	3441	GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTAAGGCAATCTCACTATT	3500
Qy	3061	TCAGAAAGCGGACCTTTTAAAGGAAAGACTAGAGATACCTTAAATATATACCGGCAATTTT	3120
Db	3501	TCAGAAAGCGGACCTTTTAAAGGAAAGACTAGAGATACCTTAAATATATACCGGCAATTTT	3560
Qy	3121	ACCAATATGCGACTGCGCAAAATTAATAACAAAGGAGTGGTAAACCTTTGGCAATGTT	3180
Db	3561	ACCAATATGCGACTGCGCAAAATTAATAACAAAGGAGTGGTAAACCTTTGGCAATGTT	3620
Qy	3181	ACCAATATGCGACTGCGCAAAATTAATAACAAAGGAGTGGTAAACCTTTGGCAATGTT	3240
Db	3621	ACCAATATGCGACTGCGCAAAATTAATAACAAAGGAGTGGTAAACCTTTGGCAATGTT	3680
Qy	3241	GGCGGAGATTAATTAACAAAGGAGTGGTAAATTTTACAGACAGTAATTAATGATGCT	3300
Db	3681	GGCGGAGATTAATTAACAAAGGAGTGGTAAATTTTACAGACAGTAATTAATGATGCT	3740
Qy	3301	GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTCCGAT	3360
Db	3741	GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTCCGAT	3800
Qy	3361	AAAATTAATATACCAAAACAGATAACAATCAAAAAGGGTATTGATGGAGAGGACTAGT	3420

```

Db 3801 AAAATTAAATATCACCACAGATAACAATCAAAAGGGTATTGATGGAGAGACTCTAGT 3860
QY 3421 TCAGATGCGCAGCAAGTAATGCCAACCTTAACATTAATAACCAAGAAATTAATAATGACAGAA 3480
Db 3861 TCAGATGCGCAGCAAGTAATGCCAACCTTAACATTAATAACCAAGAAATTAATAATGACAGAA 3920
QY 3481 GACCTAAGTATTTTCAAGTTTCAATAAAGCAGAGATTACAGCCAAAGATGAGAGATTTA 3540
Db 3921 GACCTAAGTATTTTCAAGTTTCAATAAAGCAGAGATTACAGCCAAAGATGAGAGATTTA 3980
QY 3541 ACTATTGCAACACTAATGACGTAACAGCGGTGCGAAGCCAAAGCAAGTAATTTAATAC 3600
Db 3981 ACTATTGCAACACTAATGACGTAACAGCGGTGCGAAGCCAAAGCAAGTAATTTAATAC 4040
QY 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGGTCAACAATGTGACACTAAATAGCAAAAGT 3660
Db 4041 AATGTTAAAGATTCAAAATCTCTGCTGACGGGTCAACAATGTGACACTAAATAGCAAAAGT 4100
QY 3661 AAACATCTAGCACAATGGCGAGCTGAAGCAATAGCGACAGATACCGGCTTAAC 3720
Db 4101 AAACATCTAGCACAATGGCGAGCTGAAGCAATAGCGACAGATACCGGCTTAAC 4160
QY 3721 ATTACTGCAAAATGTAGAGTAACAAAGATTAATTAATCTCTCAAAACAGTAATATC 3780
Db 4161 ATTACTGCAAAATGTAGAGTAACAAAGATTAATTAATCTCTCAAAACAGTAATATC 4220
QY 3781 ACCGGTCGAAAAGGTTACCACCAAGAGCTGACCAATTAACCAAGAGTAATATC 3840
Db 4221 ACCGGTCGAAAAGGTTACCACCAAGAGCTGACCAATTAACCAAGAGTAATATC 4280
QY 3841 GCAAGTATTACAAACCAAGAGTATACAGCGGTACGATTTCCGGTAAACAGGTAAGT 3900
Db 4281 GCAAGTATTACAAACCAAGAGTATACAGCGGTACGATTTCCGGTAAACAGGTAAGT 4340
QY 3901 GTTAGCGGACTGGTATTAACCACTAAATCCGGCTCAAAATTAACGAGAAATCCGGT 3960
Db 4341 GTTAGCGGACTGGTATTAACCACTAAATCCGGCTCAAAATTAACGAGAAATCCGGT 4400
QY 3961 GAGCTAATGTAACCAAGTGAACAGGTAACATTTGGCGGTACAATTTCCGGTAAACAGGTA 4020
Db 4401 GAGCTAATGTAACCAAGTGAACAGGTAACATTTGGCGGTACAATTTCCGGTAAACAGGTA 4460
QY 4021 AATGTTACGCAAAAGCTGGCGATTTACAGTTGGGAATGGCGAGAAATTAATGCGACA 4080
Db 4461 AATGTTACGCAAAAGCTGGCGATTTACAGTTGGGAATGGCGAGAAATTAATGCGACA 4520
QY 4081 GAAGGAGTGAACCTTTAAGCGCAAGGGAATACCTTGACTACTGAGCGGTTCTAGC 4140
Db 4521 GAAGGAGTGAACCTTTAAGCGCAAGGGAATACCTTGACTACTGAGCGGTTCTAGC 4580
QY 4141 ATCACTTCAACTAAGGTCAGGTAGACCTTTGGCTCAGAAATGGTAGCGGAAGC 4200
Db 4581 ATCACTTCAACTAAGGTCAGGTAGACCTTTGGCTCAGAAATGGTAGCGGAAGC 4640
QY 4201 ATTAATGCTGATATGACATTAATACTACAGGCACTTAACCCAGCTGGCAGGCTCG 4260
Db 4641 ATTAATGCTGATATGACATTAATACTACAGGCACTTAACCCAGCTGGCAGGCTCG 4700
QY 4261 GATATTAAGCAACCAAGCGGACCTTTGGTTAATGCGAAAGATGCTAAGCTAATGGT 4320
Db 4701 GATATTAAGCAACCAAGCGGACCTTTGGTTAATGCGAAAGATGCTAAGCTAATGGT 4760
QY 4321 GATGATCAGGTCATATGACAGAGTGAATGCAAGTCAAGCAAGCGGCTCTGATAGTG 4380
Db 4761 GATGATCAGGTCATATGACAGAGTGAATGCAAGTCAAGCAAGCGGCTCTGATAGTG 4820
QY 4381 ACTGCGCAACCTTCAAGCAGTGTCAATATCACTGGGATTTAAACACAGTAATGGGTTA 4440
Db 4821 ACTGCGCAACCTTCAAGCAGTGTCAATATCACTGGGATTTAAACACAGTAATGGGTTA 4880
QY 4441 AATATCAATTCGAAGATGTTAGAACACTGTGCGCTTTAAGAGGCAAGAAATTCAGGTG 4500

```

```

Db 4881 AATATCAATTTCAAGAGTGTAGAACTGTGGCTTTAGAGCAAGAAATTTAGGTG 4940
QY 4501 AATATATCCAGCGAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCAACCAAGCGTCTT 4560
Db 4941 AATATATCCAGCGAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCAACCAAGCGTCTT 5000
QY 4561 GAAAAAGTAAAGATTTATCTGATCAAGAAAGAAACATTTAGCTAAACTTTGGTGAAGT 4620
Db 5001 GAAAAAGTAAAGATTTATCTGATCAAGAAAGAAACATTTAGCTAAACTTTGGTGAAGT 5060
QY 4621 GCTGTACGTTTGTGTGAGCCAAATTAACAATTTACAGTCAATACACAAATGAATTTACA 4680
Db 5061 GCTGTACGTTTGTGTGAGCCAAATTAACAATTTACAGTCAATACACAAATGAATTTACA 5120
QY 4681 ACCAGCCGTCAGTCAAGTCAAGTGAATTTCTGAAGSTAAAGCGGTGTTCTCAAGTGGTAAT 4740
Db 5121 ACCAGCCGTCAGTCAAGTCAAGTGAATTTCTGAAGSTAAAGCGGTGTTCTCAAGTGGTAAT 5180
QY 4741 GCGCAGCAGTATGTACCAATTTGCTGACGATGGACACGCGTAGTCAGTAATTTGACAAG 4800
Db 5181 GCGCAGCAGTATGTACCAATTTGCTGACGATGGACACGCGTAGTCAGTAATTTGACAAG 5240
QY 4801 GTAGATTTATCCTGCAATGAAGTCAATTTATTTTCGATTAATTTACTGTGGTGTAAA 4860
Db 5241 GTAGATTTATCCTGCAATGAAGTCAATTTATTTTCGATTAATTTACTGTGGTGTAAA 5300
QY 4861 GTTCAGTACGGCTTTACCCATCTTTGTAATAAATTTACGGAGATTAACAATAAATTTT 4920
Db 5301 GTTCAGTACGGCTTTACCCATCTTTGTAATAAATTTACGGAGATTAACAATAAATTTT 5360
QY 4921 AACAGGTTATTAATG 4937
Db 5361 AACAGGTTATTAATG 5377

```

```

RESULT 10
ARI45901
LOCUS
DEFINITION Sequence 6 from patent us 6218141.
ACCESSION ARI45901
VERSION ARI45901.1 GI:15109090
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Barenkamp, S.J.
TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 6218141-A 6 17-APR-2001;
FEATURES
source location/Qualifiers
BASE COUNT 3134 a 1829 c 1832 g 2528 t
ORIGIN
Query Match 97.6%; Score 4820.2; DB 6; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGACAAACAACAAATTTACAA 60
Db 441 TAAATATACAAGATAATAAATAAATAAATCAAGATTTTGTGATGACAAACAACAAATTTACAA 500
QY 61 CACCTTTTTCGAGCTCTATATGCAAAATATTTAAAAAATAGTATATAATCCGCCATATAA 120
Db 501 CACCTTTTTCGAGCTCTATATGCAAAATATTTAAAAAATAGTATATAATCCGCCATATAA 560
QY 121 AATGCTATATCTTTCATCTTTCATCTTTAATCTTTTCATCTTTTCATCTTTTCATCTTT 180
Db 561 AATGCTATATCTTTCATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 620
QY 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTT 240

```


Qy	2401	AGCAATAGCAAAAGCCTTTAAACAACACAGTATAGAAGCTCTCGACGGGTGAATTTTACGGC	2460
Db	2841	AGCAATAGCAAAAGCCTTTAAACAACACAGTATAGAAGCTCTCGACGGGTGAATTTTACGGC	2900
Qy	2461	GTAATGGCAACATGTCATTCAATCTCAAGAAGGAGCGAAAGTTAAATTTCAAAATTTAAAA	2520
Db	2901	GTAATGGCAACATGTCATTCAATCTCAAGAAGGAGCGAAAGTTAAATTTCAAAATTTAAAA	2960
Qy	2521	CCAAACGGAACATGAACAACAAGCAAAACCTTTTACCAATTCGGTTTTTTAGCCAATATCA	2580
Db	2961	CCAAACGGAACATGAACAACAAGCAAAACCTTTTACCAATTCGGTTTTTTAGCCAATATCA	3020
Qy	2581	GCCACTGGTGGGGCTCTGTTTTTTTTTGATATATATGCCAACCAATCTCGCAGAGGGCT	2640
Db	3021	GCCACTGGTGGGGCTCTGTTTTTTTTTGATATATATGCCAACCAATCTCGCAGAGGGCT	3080
Qy	2641	GAGTTAAAAATGAGTGAATAATTAATCTCTAACGGCGCTAAATTTACCTTTAAATTCCCAT	2700
Db	3081	GAGTTAAAAATGAGTGAATAATTAATCTCTAACGGCGCTAAATTTACCTTTAAATTCCCAT	3140
Qy	2701	GTTCTGGGGGATGAGCGCTTTTAAATCAACAAGAAGCTTTAACCATTAATGCAACCAATTC	2760
Db	3141	GTTCTGGGGGATGAGCGCTTTTAAATCAACAAGAAGCTTTAACCATTAATGCAACCAATTC	3200
Qy	2761	AAITTCAGCCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC	2820
Db	3201	AAITTCAGCCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC	3260
Qy	2821	AAITTCACCTTACAACATATCCATTCCTGGGGCGTAAATGTCACCCCTTGGTGCAAAAAC	2880
Db	3261	AAITTCACCTTACAACATATCCATTCCTGGGGCGTAAATGTCACCCCTTGGTGCAAAAAC	3320
Qy	2881	AGCAGCAGCAATTACGGGAATATTAATCTCGAGAAGACGACAAATGTTACGCTAGAGGC	2940
Db	3321	AGCAGCAGCAATTACGGGGAAATATTAATCTCGAGAAGACGACAAATGTTACGCTAGAGGC	3380
Qy	2941	AAATAACGCCCTTAATCAGCAAAACATTAAGGGATAGAGTTATAAACCTTGGCAGCTTGCT	3000
Db	3381	AAATAACGCCCTTAATCAGCAAAACATTAAGGGATAGAGTTATAAACCTTGGCAGCTTGCT	3440
Qy	3001	GTTAATGGGAGTTTAAGTTTAATCTGGCAAAATCGAGATATTAAGGCAATCTCACTATT	3060
Db	3441	GTTAATGGGAGTTTAAGTTTAATCTGGCAAAATCGAGATATTAAGGCAATCTCACTATT	3500
Qy	3061	TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCCCTAAATATACCGGCAATTTT	3120
Db	3501	TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCCCTAAATATACCGGCAATTTT	3560
Qy	3121	ACCAATATGGCACTCGGAAATTAATTAACACAGAGTGGTAAACCTTGGCAATGTT	3180
Db	3561	ACCAATATGGCACTCGGAAATTAATTAACACAGAGTGGTAAACCTTGGCAATGTT	3620
Qy	3181	ACCAATGATCGCTATTAAACATTAACCTACGCTTAACGCAACCAAGAAAGCATCATC	3240
Db	3621	ACCAATGATCGCTATTAAACATTAACCTACGCTTAACGCAACCAAGAAAGCATCATC	3680
Qy	3241	GGCGGAGATATAACAACAAAAAGAAAGCTTTAAATATTAACAGACAGTAATAATGATGCT	3300
Db	3681	GGCGGAGATATAACAACAAAAAGAAAGCTTTAAATATTAACAGACAGTAATAATGATGCT	3740
Qy	3301	GAATCCAAATTCGGCGCAATATCTCGAAAAAGAGCAACCTTCAGANTTTCTCCGAT	3360
Db	3741	GAATCCAAATTCGGCGCAATATCTCGAAAAAGAGCAACCTTCAGANTTTCTCCGAT	3800
Qy	3361	AAAAATTAATATCAACCAACAGATAACATTCAAAAAGGGTATTGATGGAGAGACTCTAGT	3420
Db	3801	AAAAATTAATATCAACCAACAGATAACATTCAAAAAGGGTATTGATGGAGAGACTCTAGT	3860
Qy	3421	TCAGATCGGACAGTAATGCCAACCTTAAGTAAACCAAGAAATTTGAATTTGACAGAA	3480
Db	3861	TCAGATCGGACAGTAATGCCAACCTTAAGTAAACCAAGAAATTTGAATTTGACAGAA	3920
Qy	3481	GACCTAAGTATTTTCAGGTTTTCAATAAGCAGAGATTTACAGCAAGAGATGGTAGAGTTTA	3540

Db	3321		GACCTAAGTATTTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA	3360
Qy	3341		ACTATTGGCAACAGTAATGACCGGTAAACAGCGTCCGAAGCCAAAACAGTAACCTTTTAAC	3600
Db	3381		ACTATTGGCAACAGTAATGACCGTAACAGCGTCCGAAGCCAAAACAGTAACCTTTTAAC	4040
Qy	3501		AATGTTAAAGATTCAAAAATCTCTGCTGACGCTACAAATGTGACACTAAATAGCAAAAGT	3560
Db	4041		AATGTTAAAGATTCAAAAATCTCTGCTGACGCTACAAATGTGACACTAAATAGCAAAAGT	4100
Qy	3661		AAACATCTAGCAGCAATGGCGAGCTGAAGCAATACGCACAACGATACCGCGCTTAACCT	3720
Db	4101		AAACATCTAGCAGCAATGGCGAGCTGAAGCAATACGCACAACGATACCGCGCTTAACCT	4160
Qy	3721		ATTACTGCAAAAATGTAGAAGTAAACAAGATATTACTTCTCTCAAAACAGTAATAATATC	3780
Db	4161		ATTACTGCAAAAATGTAGAAGTAAACAAGATATTACTTCTCTCAAAACAGTAATAATATC	4220
Qy	3781		ACCGCGTCGAAAAGGTTACCAACCACAGCGCTCGACCATTAACGCCAACAAATGGCAAA	3840
Db	4221		ACCGCGTCGAAAAGGTTACCAACCACAGCGCTCGACCATTAACGCCAACAAATGGCAAA	4280
Qy	3841		GCAAGTATTACAACCAAAAACAGTGATATCAGCGGTAGCATTTCCGGTAAACACGGTAAGT	3900
Db	4281		GCAAGTATTACAACCAAAAACAGTGATATCAGCGGTAGCATTTCCGGTAAACACGGTAAGT	4340
Qy	3901		GTTAGCGCGACTGGTGATTTAACCACCTAANTCCGGCTCAAAAATGAAGCGAAATCGGCT	3960
Db	4341		GTTAGCGCGACTGGTGATTTAACCACCTAANTCCGGCTCAAAAATGAAGCGAAATCGGCT	4000
Qy	3961		GAGGCTAATGTACAAGTGCACAGAGTACAATTGGCGGTACAAATTTCCGGTAAATACGGTA	4020
Db	4401		GAGGCTAATGTACAAGTGCACAGAGTACAATTGGCGGTACAAATTTCCGGTAAATACGGTA	4460
Qy	4021		AATGTTAAGCCAAACCGTGGCGATTTAACAGTTGGGAATGGCGCAGAAATTAATGCGACA	4080
Db	4461		AATGTTAAGCCAAACCGTGGCGATTTAACAGTTGGGAATGGCGCAGAAATTAATGCGACA	4520
Qy	4081		GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC	4140
Db	4521		GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC	4580
Qy	4141		ATCACTTCAACTAAGGTCAGGTAGACCTCTTTGGCTCAGAAATGGTAGCATCCGAGGAAC	4200
Db	4581		ATCACTTCAACTAAGGTCAGGTAGACCTCTTTGGCTCAGAAATGGTAGCATCCGAGGAAC	4640
Qy	4201		ATTAAATGCTCTAATGTGACATTAATATCTACAGGCACCTTAACCCACCGTGGCAGGCTCG	4260
Db	4641		ATTAAATGCTCTAATGTGACATTAATATCTACAGGCACCTTAACCCACCGTGGCAGGCTCG	4700
Qy	4261		GATATTAAGCAACACCGCGACCTTTGGTTATTAAACGAAAAGATGCTTAAGCTAAATGGT	4320
Db	4701		GATATTAAGCAACACCGCGACCTTTGGTTATTAAACGAAAAGATGCTTAAGCTAAATGGT	4760
Qy	4321		GATGCATCAGTGATAGTACAGAAGTGAATGCAGTCAACGCAGCGGCTCTGGTAGTGTCG	4380
Db	4761		GATGCATCAGTGATAGTACAGAAGTGAATGCAGTCAACGCAGCTGGGGATTTGGTAGTGTCG	4820
Qy	4381		ACTCGGCAACCTCAAGCAGTGGAAATATCACTGGGGATTTAAACACAGTAATATGGGTTA	4440
Db	4821		ACTCGGCAACCTCAAGCAGTGGAAATATCACTGGGGATTTAAACACAGTAATATGGGTTA	4880
Qy	4441		AATATCATTTGAAAGATGGTAGAACACTGTGGCTTTAAGAGCGAAGAAATTCAGGTG	4500
Db	4881		AATATCATTTGAAAGATGGTAGAACACTGTGGCTTTAAGAGCGAAGAAATTCAGGTG	4940
Qy	4501		AAATATATCCAGCCAGGTGAGCAAGTGTAGAAGAGTAAATGAAGCGAAACCGCTCCTT	4560
Db	4941		AAATATATCCAGCCAGGTGAGCAAGTGTAGAAGAGTAAATGAAGCGAAACCGCTCCTT	5000
Qy	4561		GA AAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTTAACCTTGGTGAAGT	4620


```

Db 5001 GAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAGCTAAACTTGGTGTAGT 5060
Qy 4621 GCTGTAGCTTTGCTGAGCAAAATACAAATACAAATACAAATACAAATACAAATACAA 4680
Db 5061 GCTGTAGCTTTGCTGAGCAAAATACAAATACAAATACAAATACAAATACAAATACAA 5120
Qy 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAGCGTGTCTCAAGTGTGAAT 4740
Db 5121 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAGCGTGTCTCAAGTGTGAAT 5180
Qy 4741 GGCGCAGGATGATGACCAATGTTGCTGACGATGGACGCGTAGTCAGTAAATGGACAAG 4800
Db 5181 GGCGCAGGATGATGACCAATGTTGCTGACGATGGACGCGTAGTCAGTAAATGGACAAG 5240
Qy 4801 GTAGATTTTCATCCGCAATGAAGTCAATTTATTTTCTGATTTATTTTCTGTTGTTTAA 4860
Db 5241 GTAGATTTTCATCCGCAATGAAGTCAATTTATTTTCTGATTTATTTTCTGTTGTTTAA 5300
Qy 4861 GTTCAGTACGGGCTTTACCCATCTTGTAAATAATTCAGGAGATACAAATAAAGTATTTT 4920
Db 5301 GTTCAGTACGGGCTTTACCCATCTTGTAAATAATTCAGGAGATACAAATAAAGTATTTT 5360
Qy 4921 AACAGGTTATTATTATG 4937
Db 5361 AACAGGTTATTATTATG 5377

RESULT 11
LOCUS 125184 9323 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 6 from patent US 5549897.
ACCESSION 125184
VERSION 125184.1 GI:1605054
KEYWORDS Unknwn.
SOURCE Unknwn.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 9323)
AUTHORS Barenkamp,S.J. and St. Gene,J.W. III.
TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 5549897-A 6 27-AUG-1996;
FEATURES
    Location/Qualifiers
        1..9323
            /organism="unknown"
BASE COUNT 3134 a 1829 c 1832 g 2528 t
ORIGIN
    Query Match 97.6%; Score 4820.2; DB 6; Length 9323;
    Best Local Similarity 98.5%; Pred. No. 0;
    Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 60
Db 441 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 500
Qy 61 CACCTTTTGTGAGTCTATATGCAAAATATTTTAAATAAATAGTATAAATCCGCCATAFAA 120
Db 501 CACCTTTTGTGAGTCTATATGCAAAATATTTTAAATAAATAGTATAAATCCGCCATAFAA 560
Qy 121 AATGTTAATCTTTCATCTTTTCATCTTTAATCTTTTCATCTTTTCATCTTTTCATCTT 180
Db 561 AATGTTAATCTTTCATCTTTTCATCTTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620
Qy 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680
Qy 241 CACATGAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 681 CACATGAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
Qy 301 GACGCAATATGAATGAATTAATTTGTTCAACTAACCTTAGGAGAAATATGAACAAG 360

```

```

Db 741 GAACGCAAAATGATAAAGTAATTTTAAATTTGTTCAACTTAACCTTTAGGAGAAAAATATGAACAAG 800
Qy 361 ATATATCGTCTCAAAATTCAGCAAAACCGCTGAATGCTTTGGTGTGCTGTCTGCTGAATTTGGCA 420
Db 801 ATATATCGTCTCAAAATTCAGCAAAACCGCTGAATGCTTTGGTGTGCTGTCTGCTGAATTTGGCA 860
Qy 421 CGGGGTGTGACCAATTCACAGAAAAAGGCTTCCCGCTATGTTTACCTATCTTTTAGTGTGTAAAC 480
Db 861 CGGGGTGTGACCAATTCACAGAAAAAGGCGAGCAAAAACCTGCTCGCATGAAAGTGGCT 920
Qy 481 CACTTAGCTTTAAAGCCACTTTCCGCTATGTTTACTATCTTTTAGTGTGTAAACATCTATTCCA 540
Db 921 CACTTAGCTTTAAAGCCACTTTCCGCTATGTTTACTATCTTTTAGTGTGTAAACATCTATTCCA 980
Qy 541 CAATCTGTTTAAAGCAAGCGCTTACAAAGGAATGGATGTAGTACACGGCACACGCCACTATG 600
Db 981 CAATCTGTTTAAAGCAAGCGCAATTTAAACATCGACCAAAATGAAATGGTGTGCTGCTTTTAC 1040
Qy 601 CAAGTAGATGTTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGGAAA 660
Db 1041 AAGAAACCAAGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGGAAA 1100
Qy 661 CAATTTACATCGACCAAAATGAAATGGTGTGCAAGTTTTCACAGAAAAACAACACTCCGCC 720
Db 1101 CAATTTACATCGACCAAAATGAAATGGTGTGCAAGTTTTCACAGAAAAACAACACTCCGCC 1160
Qy 721 GTATTCACCGGTTTACATCTAACCAAAATCTCCCAATTTAAAGGGGATTTTAGATCTTAAC 780
Db 1161 GTATTCACCGGTTTACATCTAACCAAAATCTCCCAATTTAAAGGGGATTTTAGATCTTAAC 1220
Qy 781 GGAAGTCTTTTAAATCAACCAAAATGATATCAATAGGTAAGACGCAATTTATTAAAC 840
Db 1221 GGAAGTCTTTTAAATCAACCAAAATGATATCAATAGGTAAGACGCAATTTATTAAAC 1280
Qy 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTTAAGCAAAACATCAAGGGCGCTAAT 900
Db 1281 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTTAAGCAAAACATCAAGGGCGCTAAT 1340
Qy 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTAATCAGCTTTAAT 960
Db 1341 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTAATCAGCTTTAAT 1400
Qy 961 ACTGTCGGTAAAGACGCGAGTGAATCTTATTTGGTGGCAAAAGTGAAGAAACGAGGTGTG 1020
Db 1401 ACTGTCGGTAAAGACGCGAGTGAATCTTATTTGGTGGCAAAAGTGAAGAAACGAGGTGTG 1460
Qy 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTCGAGGGCAAAAATCAGCATCAGGAT 1080
Db 1461 ATTAGCGTAAATGGTGGCAGCATTTCTTACTCGAGGGCAAAAATCAGCATCAGGAT 1520
Qy 1081 ATAATAACCCCAACCATTTACTTACAGCATTTGCGCGCTGAAATGAGCGGCTCAATCTG 1140
Db 1521 ATAATAACCCCAACCATTTACTTACAGCATTTGCGCGCTGAAATGAGCGGCTCAATCTG 1580
Qy 1141 GGCATATTTTTCGCAAGCGGTAAACATTAATGTCGCTGCTGCCACTATTGCGAACCA 1200
Db 1581 GGCATATTTTTCGCAAGCGGTAAACATTAATGTCGCTGCTGCCACTATTGCGAACCA 1640
Qy 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAAAGTGAAGCGCATATTGTTCTTCGCC 1260
Db 1641 GGTAAACTTTCTGCTGATTTCTGTAAGCAAAAGTGAAGCGCATATTGTTCTTCGCC 1700
Qy 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTGAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
Db 1701 AAAGAGGTGAAGCGGAAATTTGGCGGTGAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1760
Qy 1321 GGCAAGCTGATGATTAACAGCGGTAATTTAAAGTCAATTAACAGCTGAGTATCGACCTT 1380
Db 1761 GGCAAGCTGATGATTAACAGCTGATTAAGTCCGATTAAGTCAATTAACAGCTGAGTATCGACCTT 1820
Qy 1381 TCAGGTAAGAGGGGGAGAAACCTTACCTGGCGCTGACGAGCGCGGAGGTGAAAC 1440
Db 1821 TCAGGTAAGAGGGGGAGAAACCTTACCTGGCGCTGACGAGCGCGGAGGTGAAAC 1880

```


us-10-092-880-3.rge

Tue Mar 25 08:39:00 2003

4041 AATGTTAAAGATTCACAAATCTCTGCTGAGGTCAATGTGACATTAATACGAAGTG 4100
 3661 AAAACATCTAGCAGCAATGGCGAGCTGAAGCAATACGCAACAGATACGGCTTAAC 3720
 4101 AAAACATCTAGCAGCAATGGCGAGCTGAAGCAATACGCAACAGATACGGCTTAAC 4160
 3721 ATTACTGCAAAATGTAGAGTAACCAAGATATTAATCTCTCAAAACAGATTAATATC 3780
 4161 ATTACTGCAAAATGTAGAGTAACCAAGATATTAATCTCTCAAAACAGATTAATATC 4220
 3781 ACCGGCTCGGAAAGGTTACACACAGAGCTCGACCATTAAGCCAAACAAATGGCAA 3840
 4221 ACCGGCTCGGAAAGGTTACACACAGAGCTCGACCATTAAGCCAAACAAATGGCAA 4280
 3841 GCAAGTATTACAAACAAACAGAGTATACAGGGTACGATTTCCGGTAAACAGGTAAGT 3900
 4281 GCAAGTATTACAAACAAACAGAGTATACAGGGTACGATTTCCGGTAAACAGGTAAGT 4340
 3901 GTTAGCGGACCTGGTGAATTAACCACTAAATCCGGCTCAAAATTTGAAGCAAAATCGGT 3960
 4341 GTTAGCGGACCTGGTGAATTAACCACTAAATCCGGCTCAAAATTTGAAGCAAAATCGGT 4400
 3961 GAGGCTAATGTAAACAGTGAACAGGTACAATTTGGCGGTACAAATTTCCGGTAAACAGGTA 4020
 4401 GAGGCTAATGTAAACAGTGAACAGGTACAATTTGGCGGTACAAATTTCCGGTAAACAGGTA 4460
 4021 AATGTTACGCAACAGCTGGCGATTTAAACAGTTGGGAATGGCGCAAGAAATTAATGCCACA 4080
 4461 AATGTTACGCAACAGCTGGCGATTTAAACAGTTGGGAATGGCGCAAGAAATTAATGCCACA 4520
 4081 GAAAGAGCTCAACCTTAACCGCAACAGGGAATACCTGACCTACGAGCGGCTTACG 4140
 4521 GAAAGAGCTCAACCTTAACCGCAACAGGGAATACCTGACCTACGAGCGGCTTACG 4580
 4141 ATCACTTCACTAAGGTCAGGTAGAGCTCTTGCTGCAAGTGAATGATTCAGCAAGAGC 4200
 4581 ATCACTTCACTAAGGTCAGGTAGAGCTCTTGCTGCAAGTGAATGATTCAGCAAGAGC 4640
 4201 ATTAATGCTGATATGTCAGATTAATTAATACAGGCACTTAACACCGTGGCGAGCTCG 4260
 4641 ATTAATGCTGATATGTCAGATTAATTAATACAGGCACTTAACACCGTGGCGAGCTCG 4700
 4261 GATATTAAAGCAACAGCGGCACTTGGTATTAAACGCAAGAGATGTAAGCTAAATGGT 4320
 4701 GATATTAAAGCAACAGCGGCACTTGGTATTAAACGCAAGAGATGTAAGCTAAATGGT 4760
 4321 GATGCAATCAGGTGATAGTACAGAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4380
 4761 GATGCAATCAGGTGATAGTACAGAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4820
 4381 ACTCGGCAACCTCAAGCAGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440
 4821 ACTCGGCAACCTCAAGCAGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4880
 4441 AATATCATTTGGAAGATGGTAGAAGTGTGCGCTTAAGAGCAAGGAAATTAAGGTTG 4500
 4881 AATATCATTTGGAAGATGGTAGAAGTGTGCGCTTAAGAGCAAGGAAATTAAGGTTG 4940
 4501 AATATATATCCAGCCAGGTTAGCAAGTGTAGAGAGTGAATTAAGAGCAAGGAAATTAAG 4560
 4941 AATATATATCCAGCCAGGTTAGCAAGTGTAGAGAGTGAATTAAGAGCAAGGAAATTAAG 5000
 4561 GAAAGAGTAAAGATTTATCTGATGAAGAAAGAGAAATAGCTAACTTGGTGAAGT 4620
 5001 GAAAGAGTAAAGATTTATCTGATGAAGAAAGAGAAATAGCTAACTTGGTGAAGT 5060
 4621 GCTGTAGCTTTGTTGAGCCAAATTAATCAATTAAGTCAATACACAAATTAATTAACA 4680
 5061 GCTGTAGCTTTGTTGAGCCAAATTAATCAATTAAGTCAATACACAAATTAATTAACA 5120
 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTGAAGGCTGTTCTCAAGTGAAT 4740

5121 ACCAGACCGTCAAGTCAAGTGAATATTTCTGAAGGTAGGCGTGTCTTCAAGTGTGAAT 5180
 4741 GGCACAGATATGTACCAATGTGCTGAGATGGACAGCGGTAGTCAATATGACAAAG 4800
 5181 GGCACAGATATGTACCAATGTGCTGAGATGGACAGCGGTAGTCAATATGACAAAG 5240
 4801 GTAGATTTTATCCTCCGCAATGAAGTCAATTTATTTTCTGATTTATTTACTGTGGGTAAA 4860
 5241 GTAGATTTTATCCTCCGCAATGAAGTCAATTTATTTTCTGATTTATTTACTGTGGGTAAA 5300
 4861 GTTCAAGTACGGGCTTTACCCATCTTGTAAATAATTCAGGAGAAATACAAATAAAGTATTTT 4920
 5301 GTTCAAGTACGGGCTTTACCCATCTTGTAAATAATTCAGGAGAAATACAAATAAAGTATTTT 5360
 4921 AACAGGTTATTATTATG 4937
 5361 AACAGGTTATTATTATG 5377

RESULT 12
 LOCUS I35773 9323 bp DNA linear PAT 13-MAY-1997
 DEFINITION Sequence 6 from patent US 5603938.
 ACCESSION I35773
 VERSION I35773.1 GI:2086997
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 9323)
 AUTHORS Barenkamp, S. J.
 TITLE High molecular weight surface proteins of non-typeable haemophilus
 JOURNAL Patent: US 5603938-A 6 18-FEB-1997;
 FEATURES Location/Qualifiers
 1..9323
 source /organism="unknown"

BASE COUNT 3134 a 1829 c 1832 g 2528 t

Query Match 97.6%; Score 4820.2; DB 6; Length 9323;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAAATTAACAA 60
 Db 441 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAAATTAACAA 500
 QY 61 CACCTTTTTCGAGTCTATATGCAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
 Db 501 CACCTTTTTCGAGTCTATATGCAATATTTTAAAAAATAGTATAAATCCGCCATATAA 560
 QY 121 AATGATATAATCTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 180
 Db 561 AATGATATAATCTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 620
 QY 181 CTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 240
 Db 621 CTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 680
 QY 241 CACATGAAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 Db 681 CACATGAAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
 QY 301 GAACGCAAAATGATFAAAGTAAATTAATTTGTTCACTTAACCTTAGGAGAGGAGGAGGAGGAGGAGG 360
 Db 741 GAACGCAAAATGATFAAAGTAAATTAATTTGTTCACTTAACCTTAGGAGAGGAGGAGGAGGAGGAGG 800
 QY 361 ATATATCTCTCAAAATTCAGCAAAACCGCTGAATGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 Db 801 ATATATCTCTCAAAATTCAGCAAAACCGCTGAATGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 860
 QY 421 CGGGTGTGTGACCAATTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480

Db 861 CGGGTGTGACCAATCCACAGAAAAAGGAGGAGGAAAAACCTGCTCGCATGAAAGTGGCT 920
Qy 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAAACATCTATTCCA 540
Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAAACATCTATTCCA 980
Qy 541 CAATCTGTTTACCAAGCGGCTTACAGGAATGGATGTAGTACAGCGCACAGCCACTATG 600
Db 981 CAATCTGTTTACCAAGCGGCAATTTAATCATCGACCAAAATGAATGTGTGAGTTTATC 1040
Qy 601 CAGTAGATGGTAATAAACCATTATCCGCAACAGTGTGACGTATCATTAATTTGAAA 660
Db 1041 AGAAAAACAAGTAATAAACCATTATCCGCAACAGTGTGACGTATCATTAATTTGAAA 1100
Qy 661 CAATTTAATCATCGACCAAAATGAATGGTGTGAGTTTACAGAAAAACAACCTCGGC 720
Db 1101 CAATTTAATCATCGACCAAAATGAATGGTGTGAGTTTACAGAAAAACAACCTCGGC 1160
Qy 721 GTATTCACCGTGTACATCAACCAATCTCCCAATTTAAAGGGATTTAGATTCTAAC 780
Db 1161 GTATTCACCGTGTACATCAACCAATCTCCCAATTTAAAGGGATTTAGATTCTAAC 1220
Qy 781 GGCAAGTCTTTTAAATCAACCAATGTATCAATAGGTAAGACGCAATTTAATTAAC 840
Db 1221 GGCAAGTCTTTTAAATCAACCAATGTATCAATAGGTAAGACGCAATTTAATTAAC 1280
Qy 841 ACTAATGGCTTTACGGCTTACGGCTAGACATTTCTAAGCAAAACATCAAGCGCGTAAT 900
Db 1281 ACTAATGGCTTTACGGCTTACGGCTAGACATTTCTAAGCAAAACATCAAGCGCGTAAT 960
Qy 901 TTACACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAATTTGCAATCAAGGTTAAT 1020
Db 1341 TTACACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAATTTGCAATCAAGGTTAAT 1400
Qy 961 ACTGTCGTTAAAGCGGAGTGAATCTTATGTTGGCAAGTGAAGGAGGAGGTTG 1460
Db 1401 ACTGTCGTTAAAGCGGAGTGAATCTTATGTTGGCAAGTGAAGGAGGAGGTTG 1520
Qy 1021 ATTAGCGTAAATGTPGCGAGCATTTCTTACTCGCAGGCAAAATCAACATCAGCGAT 1080
Db 1461 ATTAGCGTAAATGTPGCGAGCATTTCTTACTCGCAGGCAAAATCAACATCAGCGAT 1140
Qy 1081 ATAATAACCCAAACCATTTACTTACAGATTTGCGCGCTGAAATGAAGCGTCAATCTG 1180
Db 1521 ATAATAACCCAAACCATTTACTTACAGATTTGCGCGCTGAAATGAAGCGTCAATCTG 1240
Qy 1141 GCGGATATTTTGCAGAGCGGTAAATTAATGTCGCTGCTGCCACTATTCGAAACCA 1300
Db 1581 GCGGATATTTTGCAGAGCGGTAAATTAATGTCGCTGCTGCCACTATTCGAAACCA 1360
Qy 1201 GGTAAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCGGCAATTAATTTTCCGCC 1420
Db 1641 GGTAAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCGGCAATTAATTTTCCGCC 1480
Qy 1261 AAAGAGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAACAGCTAAAGG 1540
Db 1701 AAAGAGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAACAGCTAAAGG 1600
Qy 1321 GCGAAGCTGATGATTAAGCGGATAAAGTCAATTAATAACAGGTGCAAGTATCGACCT 1660
Db 1761 GCGAAGCTGATGATTAAGCTGATTAAGTCAATTAATAACAGGTGCAAGTATCGACCT 1720
Qy 1381 TCAGTAAGAGGAGGAGAACTTACCTTGGGGTGACGAGCGCGGCAAGGTAAAC 1780
Db 1821 TCAGTAAGAGGAGGAGAACTTACCTTGGGGTGACGAGCGCGGCAAGGTAAAC 1840
Qy 1441 GGCATTCATTAAGCAAAACCTCTTTAGAAAAAGGTCAACCATCAATGATCAGG 1900
Db 1881 GGCATTCATTAAGCAAAACCTCTTTAGAAAAAGGTCAACCATCAATGATCAGG 1960
Qy 1501 AAAGAAAAAGGAGCGGCTATTTGTTGGGCGATATGCGGTTAATGACGCAATAT 2000
Db 1941 AAAGAAAAAGGAGCGGCTATTTGTTGGGCGATATTTGCGTTAATGACGCAATAT 2060

Qy 1561 AACGCTCAAGGTAGTGGTATTCGCTAAAACCGGTGGTGTGTTGAGACATCGGGGAT 1620
Db 2001 AACGCTCAAGGTAGTGGTATTCGCTAAAACCGGTGGTGTGTTGAGACATCGGGGAT 2060
Qy 1621 TATTTATCCATTCACAGCAATGCAATGTTTAAACAAAAGAGTGGTGTAGACCCCTGAT 1680
Db 2061 TATTTATCCATTCACAGCAATGCAATGTTTAAACAAAAGAGTGGTGTAGACCCCTGAT 2120
Qy 1681 GATGTAACAATTTGAAGCCGAGAGACCCCTTCGCAATTAATACCGGTATAATGATGATTC 1740
Db 2121 GATGTAACAATTTGAAGCCGAGAGACCCCTTCGCAATTAATACCGGTATAATGATGATTC 2180
Qy 1741 CCAACAGGACCGGTGAAGCAAGCGACCTTAAAAAAAATAGCAACTCAAAACAGGCTA 1800
Db 2181 CCAACAGGACCGGTGAAGCAAGCGACCTTAAAAAAAATAGCAACTCAAAACAGGCTA 2240
Qy 1801 ACCAATACAATTTTCAAAATTTCTGAAAACGCGCTGGACAATGAATATAAGCGCATCA 1860
Db 2241 ACCAATACAATTTTCAAAATTTCTGAAAACGCGCTGGACAATGAATATAAGCGCATCA 2300
Qy 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920
Db 2301 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 2360
Qy 1921 AGTAAAGGTACGCGTGGCGAGCGGTTTCAAGATTGATGAGATATTACTTCTAAAAGCGGA 1980
Db 2361 AGTAAAGGTACGCGTGGCGAGCGGTTTCAAGATTGATGAGATATTACTTCTAAAAGCGGA 2420
Qy 1981 AATTTAAACCATTTATTTACCGCGCTTCCGTAAGTGTGTAATAAATATTACGCTTATCAG 2040
Db 2421 AATTTAAACCATTTATTTACCGCGCTTCCGTAAGTGTGTAATAAATATTACGCTTATCAG 2480
Qy 2041 GGTTTTTTAAATTTTACCGCGCTTCCGTAAGTGTGTAATAAATATTACGCTTATCAG 2100
Db 2481 GGTTTTTTAAATTTTACCGCGCTTCCGTAAGTGTGTAATAAATATTACGCTTATCAG 2160
Qy 2101 GACGCGCAAAATGCTAAATTTGTCGCCAGGCGCTGTAACCATTTACAGAGAGGAA 2540
Db 2541 GACGCGCAAAATGCTAAATTTGTCGCCAGGCGCTGTAACCATTTACAGAGAGGAA 2600
Qy 2161 GATTTACAGGCTTAAACACATCTTTAAACGGAAGGTTAAAGTCTGAATATCATTTCA 2220
Db 2601 GATTTACAGGCTTAAACACATCTTTAAACGGAAGGTTAAAGTCTGAATATCATTTCA 2660
Qy 2221 TCAGTCAATTAATTTAAACCCACATCTTACGCGCAATTAACATATCTGGGAATAACA 2280
Db 2661 TCAGTCAATTAATTTAAACCCACATCTTACGCGCAATTAACATATCTGGGAATAACA 2720
Qy 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTTGGCAAAACAGCCATGATTCGACTGG 2340
Db 2721 ATTAACCAACTACGAGAAAGAACACCTCGTATTTGGCAAAACAGCCATGATTCGACTGG 2780
Qy 2341 AACGTCAGTCTTAAATCTAGACAGCGCAATTTTACCTTTAATAATACATTTCA 2400
Db 2781 AACGTCAGTCTTAAATCTAGACAGCGCAATTTTACCTTTAATAATACATTTCA 2840
Qy 2401 AGCAATAGCAAGGCTTAAACACACATAGAGCTCTGACGGGTGAATTTAACGGC 2460
Db 2841 AGCAATAGCAAGGCTTAAACACACATAGAGCTCTGACGGGTGAATTTAACGGC 2900
Qy 2461 GTAATGGCAACATGTCATTTCAATCTCAAGAGAGGAGGCAAGTTAATTTCAAAATTA 2520
Db 2901 GTAATGGCAACATGTCATTTCAATCTCAAGAGAGGAGGCAAGTTAATTTCAAAATTA 2960
Qy 2521 CCAACAGAGAACATGAACACACACCTTTACCAATTTGCGGTGTTTACGCAATATACA 2580
Db 2961 CCAACAGAGAACATGAACACACACCTTTACCAATTTGCGGTGTTTACGCAATATACA 3020
Qy 2581 GGCACGTGGGGGCTCTGTTTTTGTATATATATGCAACCATTTGCGAGAGGGCT 2640
Db 3021 GGCACGTGGGGGCTCTGTTTTTGTATATATATGCAACCATTTGCGAGAGGGCT 3080

us-10-092-880-3.rge

Tue Mar 25 08:39:00 2003

QY	2641	GAGTTAAATGAGTGAATTAATATCTCTAAGGGGCTAAATTTACCTTAAATTCCTAT	2700
Db	3081	GAGTTAAATGAGTGAATTAATATCTCTAAGGGGCTAAATTTACCTTAAATTCCTAT	3140
QY	2701	GTTCCGGCGATGAGCTTTTAAATCAACAAAGACATTAACCAATAATGCAACCAATCA	2760
Db	3141	GTTCCGGCGATGAGCTTTTAAATCAACAAAGACATTAACCAATAATGCAACCAATCA	3200
QY	2761	AATTTGAGCTCAGACAGCAAGAGATGATTTTATGACGGGTACGCAACCAATGCGATC	2820
Db	3201	AATTTGAGCTCAGACAGCAAGAGATGATTTTATGACGGGTACGCAACCAATGCGATC	3260
QY	2821	AATTTGAGCTCAGACAGCAAGAGATGATTTTATGACGGGTACGCAACCAATGCGATC	2880
Db	3261	AATTTGAGCTCAGACAGCAAGAGATGATTTTATGACGGGTACGCAACCAATGCGATC	3320
QY	2881	AGCAGCAGCATACGGGGATATCTATCAGAGAAAGCAGCAAAATGTTACGCTAGAAGCC	2940
Db	3321	AGCAGCAGCATACGGGGATATCTATCAGAGAAAGCAGCAAAATGTTACGCTAGAAGCC	3380
QY	2941	AATTAACGGGCTTAACTGCGGAAATGCGGATAGAGTATATAAACTTTGGCAGCTTGCTC	3000
Db	3381	AATTAACGGGCTTAACTGCGGAAATGCGGATAGAGTATATAAACTTTGGCAGCTTGCTC	3440
QY	3001	GTTAATGGGAGTTTAACTGCGGAAATGCGGATAGAGTATATAAACTTTGGCAGCTTGCTC	3060
Db	3441	GTTAATGGGAGTTTAACTGCGGAAATGCGGATAGAGTATATAAACTTTGGCAGCTTGCTC	3500
QY	3061	TCAGAAAGCGGCACTTTTAAAGGAAAGCTAGAGATACCTTAATATACCGGCAATTTT	3120
Db	3501	TCAGAAAGCGGCACTTTTAAAGGAAAGCTAGAGATACCTTAATATACCGGCAATTTT	3560
QY	3121	ACCAATATGGGCACTGCGGAAATTAATATACCAAGGAGTGGTAAACTTGGCAATGTT	3180
Db	3561	ACCAATATGGGCACTGCGGAAATTAATATACCAAGGAGTGGTAAACTTGGCAATGTT	3620
QY	3181	ACCAATATGGGCACTGCGGAAATTAATATACCAAGGAGTGGTAAACTTGGCAATGTT	3240
Db	3621	ACCAATATGGGCACTGCGGAAATTAATATACCAAGGAGTGGTAAACTTGGCAATGTT	3680
QY	3241	GCGGAGATATATCAACAAAGGAAAGCTTAAATATTAACAGACATTAATGATGCT	3300
Db	3681	GCGGAGATATATCAACAAAGGAAAGCTTAAATATTAACAGACATTAATGATGCT	3740
QY	3301	GAAATCCAAATGCGGCAATATCTCGCAAAAGGAAAGCAACCTCAGATTTCTTCGAT	3360
Db	3741	GAAATCCAAATGCGGCAATATCTCGCAAAAGGAAAGCAACCTCAGATTTCTTCGAT	3800
QY	3361	AAATTAATATCAACAAAGGAAAGCTTAAATATTAACAGACATTAATGATGCT	3420
Db	3801	AAATTAATATCAACAAAGGAAAGCTTAAATATTAACAGACATTAATGATGCT	3860
QY	3421	TCAGATGCGGCAATGCGGCAATATCTCGCAAAAGGAAAGCAACCTCAGATTTCTTCGAT	3480
Db	3861	TCAGATGCGGCAATGCGGCAATATCTCGCAAAAGGAAAGCAACCTCAGATTTCTTCGAT	3920
QY	3481	GACCTAAGTATTCAGGTTTCAATTAAGCAGAGATTAACGCCAAAGGATGGTAGAGTTA	3540
Db	3921	GACCTAAGTATTCAGGTTTCAATTAAGCAGAGATTAACGCCAAAGGATGGTAGAGTTA	3980
QY	3541	ACTATTGGCAACAGTAATGAGGTAACAGCGGTGCGGCAAGGAAAGCAACCTTAACT	3600
Db	3981	ACTATTGGCAACAGTAATGAGGTAACAGCGGTGCGGCAAGGAAAGCAACCTTAACT	4040
QY	3601	AATGTTAAAGATTCAAAATCTCTGCTCAGCTCAGATGACACTTAATAGCAAGTG	3660
Db	4041	AATGTTAAAGATTCAAAATCTCTGCTCAGCTCAGATGACACTTAATAGCAAGTG	4100
QY	3661	AAAACATCTAGCAGCAATGGCGGAGTGAAGCAATAGCAGCAACGATACCGCTTAAT	3720
Db	4101	AAAACATCTAGCAGCAATGGCGGAGTGAAGCAATAGCAGCAACGATACCGCTTAAT	4160
QY	3721	ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATC	3780

Db	4161	ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATC	4220
QY	3781	ACCGCTCGGAAAGGTTTACCAACACAGCGCTGACCATTAACGCAACAAATGGCAAA	3840
Db	4221	ACCGCTCGGAAAGGTTTACCAACACAGCGCTGACCATTAACGCAACAAATGGCAAA	4280
QY	3841	GCAAGTATTACAAACAAACAGCGTATATACGCGGTAGCATTTCCGGTAAACACCGTAAGT	3900
Db	4281	GCAAGTATTACAAACAAACAGCGTATATACGCGGTAGCATTTCCGGTAAACACCGTAAGT	4340
QY	3901	GTTAGCGGACTGCTGATTTAACCACTAAATCGCGCTCAAAATTTGAAGCGAAATTCGGGT	3960
Db	4341	GTTAGCGGACTGCTGATTTAACCACTAAATCGCGCTCAAAATTTGAAGCGAAATTCGGGT	4400
QY	3961	GAGCTAATGTAAACAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAACGTA	4020
Db	4401	GAGCTAATGTAAACAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAACGTA	4460
QY	4021	AATGTTACGGCAACGCTTGGCGATTTAACAGTTGGGAATGGCGCAGAAATTAATCGACA	4080
Db	4461	AATGTTACGGCAACGCTTGGCGATTTAACAGTTGGGAATGGCGCAGAAATTAATCGACA	4520
QY	4081	GAAGGAGTGAACCTTAAACCGCAACAGGGAATACCTTGAATCTGAAGCCGCTTCTAGC	4140
Db	4521	GAAGGAGTGAACCTTAAACCGCAACAGGGAATACCTTGAATCTGAAGCCGCTTCTAGC	4580
QY	4141	ATCAGCTCAACTTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTATGATCGCAGGAGC	4200
Db	4581	ATCAGCTCAACTTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTATGATCGCAGGAGC	4640
QY	4201	ATTAATGCTCTAAATGTGACATTAATATCTACAGGACCTTTAACCCCGTGGCGAGCTCG	4260
Db	4641	ATTAATGCTCTAAATGTGACATTAATATCTACAGGACCTTTAACCCCGTGGCGAGCTCG	4700
QY	4261	GATATTAAGCAACACCGGCACTTGGTTTATTAACGCAAAAGATGCTAACTAAATGGT	4320
Db	4701	GATATTAAGCAACACCGGCACTTGGTTTATTAACGCAAAAGATGCTAACTAAATGGT	4760
QY	4321	GATGATCAGGTGATAGTACAGAGTGAATGAGTCAACGCAAGCGGCTCTGGTAGTGTG	4380
Db	4761	GATGATCAGGTGATAGTACAGAGTGAATGAGTCAACGCAAGCGGCTCTGGTAGTGTG	4820
QY	4381	ACTGGGCAACCTCAACAGCTGTAATATCAGTGGGATTTAAACACAGTAAATGGGTTA	4440
Db	4821	ACTGGGCAACCTCAACAGCTGTAATATCAGTGGGATTTAAACACAGTAAATGGGTTA	4880
QY	4441	AATATCATTTCCGAAGATGGTAGAAGACACTGCGCTTAAAGAGGCAAGGAAATTTGAGGTG	4500
Db	4881	AATATCATTTCCGAAGATGGTAGAAGACACTGCGCTTAAAGAGGCAAGGAAATTTGAGGTG	4940
QY	4501	AAATATATCCAGCCAGGTGATAGCAAGTGTAGCAAGCAATTTGAAGCGAAACGCTCTT	4560
Db	4941	AAATATATCCAGCCAGGTGATAGCAAGTGTAGCAAGCAATTTGAAGCGAAACGCTCTT	5000
QY	4561	GAAGAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAACTTTGGTGAAGT	4620
Db	5001	GAAGAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAACTTTGGTGAAGT	5060
QY	4621	GCTGTACGTTTGTGAGCCAAATTAATCAATTTACAGTCAATACACAAATTAATTTACA	4680
Db	5061	GCTGTACGTTTGTGAGCCAAATTAATCAATTTACAGTCAATACACAAATTAATTTACA	5120
QY	4681	ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAGGGCTGTTCTCAAGTGGTAAAT	4740
Db	5121	ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAGGGCTGTTCTCAAGTGGTAAAT	5180
QY	4741	GGCCGACGAGTATGTACCAATTTGCTGAGTGGACAGCCGCTAGTCAATTTGACAAAG	4800
Db	5181	GGCCGACGAGTATGTACCAATTTGCTGAGTGGACAGCCGCTAGTCAATTTGACAAAG	5240
QY	4801	GTAGATTTCACTCCATGAAGTCAATTTTATTTCTGCTATTTTACTGTGTGGTGTAAA	4860

Db	801	ATATATCGTCTCAAAATTCAGCAAAAGCGCTGAATGCTTTGGTGGCTGTCTCAAAATGGCA	860
Qy	421	CGGGTTGTGACCAATTCACAGAAAAGGCTCCGCTATGTTACTATCTTTAGGTGTAAC	480
Db	861	CGGGTTGTGACCAATTCACAGAAAAGGCGAAGAAAACCTGCTCGCATGAAGCTGGT	920
Qy	481	CACATTAGCGTTAAAGCCATTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA	540
Db	921	CACATTAGCGTTAAAGCCATTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA	980
Qy	541	CAATCTGTTTTAGCAAGCGGCTTACAAGAAATGGATGTAGTACACGGCACACCACTATG	600
Db	981	CAATCTGTTTTAGCAAGCGGCAATTTACATCGACCAAAATCAAAATGGTGCAGTTTTTAC	1040
Qy	601	CAAGTAGATGGTAAATAAACCAATATTCGCAACAGTGTGACGCTATCATTAATTTGAAA	660
Db	1041	AGAAACAAGTAATAAACCAATATTCGCAACAGTGTGACGCTATCATTAATTTGAAA	1100
Qy	661	CAATTTAATCATCGACCAAAATGAATGGTGCAGTTTTTACAGAAAACAACAACCTCGCC	720
Db	1101	CAATTTAATCATCGACCAAAATGAATGGTGCAGTTTTTACAGAAAACAACAACCTCGCC	1160
Qy	721	GTATTTCAACCGTGTACATCTAACCAAAATCTCCAAATTTAAAGGGATTTTAGATTCTAAC	780
Db	1161	GTATTTCAACCGTGTACATCTAACCAAAATCTCCAAATTTAAAGGGATTTTAGATTCTAAC	1220
Qy	781	GGACAAGTCTTTTTAATCAACCCAAATGGTATCACAAATAGTAAAGCACTATTATTAC	840
Db	1221	GGACAAGTCTTTTTAATCAACCCAAATGGTATCACAAATAGTAAAGCACTATTATTAC	1280
Qy	841	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT	900
Db	1281	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT	1340
Qy	901	TTCACTTTCAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGCAATTCAGGTTTAAT	960
Db	1341	TTCACTTTCAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGCAATTCAGGTTTAAT	1400
Qy	961	ACTGTCGGTAAAGACGGCAGTGTAAATCTTAATTTGTTGGCAAGTGAAGACGGGTGTG	1020
Db	1401	ACTGTCGGTAAAGACGGCAGTGTAAATCTTAATTTGTTGGCAAGTGAAGACGGGTGTG	1460
Qy	1021	ATTAGCGTAAATGGTGGCAGCATTTCTTTACTGCGAGGCAAAAATCACCATCAGCGAT	1080
Db	1461	ATTAGCGTAAATGGTGGCAGCATTTCTTTACTGCGAGGCAAAAATCACCATCAGCGAT	1520
Qy	1081	ATTAATAACCCAAACCATTTACTACAGCATTTCCCGCTCGAAATGAAGCGGTCAATCTG	1140
Db	1521	ATTAATAACCCAAACCATTTACTACAGCATTTCCCGCTCGAAATGAAGCGGTCAATCTG	1580
Qy	1141	GGCGATATTTTGGCAAGGGCGGTAAACATTAATGTCGGTCTGCCACTATTCGAAACCAA	1200
Db	1581	GGCGATATTTTGGCAAGGGCGGTAAACATTAATGTCGGTCTGCCACTATTCGAAACCAA	1640
Qy	1201	GGTAACTTTCTGCTGATTCGTAAAGCAAGATAAAGGGGCAATATTCGTTCTTCCGCC	1260
Db	1641	GGTAACTTTCTGCTGATTCGTAAAGCAAGATAAAGGGGCAATATTCGTTCTTCCGCC	1700
Qy	1261	AAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC	1320
Db	1701	AAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC	1380
Qy	1321	GGCAAGCTGATGATTAACGGCATTAAGTCAATTTAAACACAGGTGCAGTTATCGACCTT	1380
Db	1761	GGCAAGCTGATGATTAACGGCATTAAGTCAATTTAAACACAGGTGCAGTTATCGACCTT	1820
Qy	1381	TCAGGTAAAGAGGGGGAGAACTTTACCTTTGGCGGTGACGAGCGCGGAAGGTAAAAAC	1440
Db	1821	TCAGGTAAAGAGGGGGAGAACTTTACCTTTGGCGGTGACGAGCGCGGAAGGTAAAAAC	1880
Qy	1441	GGCAATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC	1500

Db	1881	GGCATTCAATTAGCAAAAGAAAACCTCTTTAGAAAAGGGCTCAACCAATCAATGATCAGGC	1940
Qy	1501	AAAGAAAAAGGGAGCGGCCTATTGTGTGGGGCGATATTGCGTTAATTCAGCGCAATATT	1560
Db	1941	AAAGAAAAAGGGAGCGGCCTATTGTGTGGGGCGATATTGCGTTAATTCAGCGCAATATT	2000
Qy	1561	AACGCTCAAGGTAGTGGTGATATCGCTATAAACCGGTGGTTTCTGGGAGACATCGGGCAT	1620
Db	2001	AACGCTCAAGGTAGTGGTGATATCGCTATAAACCGGTGGTTTCTGGGAGACATCGGGCAT	2060
Qy	1621	TATTTATTCATTGACAGCAATGCAATGTGTTAAACAAAGAGTGGTTCTAGACCCCTGAT	1680
Db	2061	TATTTATTCATTGACAGCAATGCAATGTGTTAAACAAAGAGTGGTTCTAGACCCCTGAT	2120
Qy	1681	GATGTAACTAATTGAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGATTC	1740
Db	2121	GATGTAACTAATTGAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGATTC	2180
Qy	1741	CCAACAGCACCGGTGAAGCAAGCAGCCCTAAAAAAATAGGGAACCTCAAAACAACGCTA	1800
Db	2181	CCAACAGCACCGGTGAAGCAAGCAGCCCTAAAAAAATAGGGAACCTCAAAACAACGCTA	2240
Qy	1801	ACCAATACAACCTATTTCAAAATTAATCTGAAAACCGCTGGACAATGAANTATACGGCATCA	1860
Db	2241	ACCAATACAACCTATTTCAAAATTAATCTGAAAACCGCTGGACAATGAANTATACGGCATCA	2300
Qy	1861	AGAAACTTACCGTTTAATAGCTCAATCAAGTCCGAAGCAACTCCCATTAATTCCTCAT	1920
Db	2301	AGAAACTTACCGTTTAATAGCTCAATCAAGTCCGAAGCAACTCCCATTAATTCCTCAT	2360
Qy	1921	AGTAPAGGTCAGCGTGGCGGAGGGGTTTCAGATTGATGGAGATATTACTTCTAAAGCGCGA	1980
Db	2361	AGTAAAGGTCAGCGTGGCGGAGGGGTTTCAGATTGATGGAGATATTACTTCTAAAGCGCGA	2040
Qy	1981	AAATTAACCAATTATTCTGGCGGATGGGTGATGTTTCATAAAAATATTACGTTGATCAG	2040
Db	2041	AAATTAACCAATTATTCTGGCGGATGGGTGATGTTTCATAAAAATATTACGTTGATCAG	2080
Qy	2041	GGTTTTTAATAATTACCGCCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAGCACGC	2100
Db	2081	GGTTTTTAATAATTACCGCCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAGCACGC	2160
Qy	2101	GACCGGCAATGCTAAAAATTGTGCGCCAGGGCACTGTAACCATTTACAGGAGAGGGAATA	2160
Db	2161	GATTTGAGGCTACACAGTATCTTTAAACGGGAAGGGTAAAGTCTGGAATPATCATTTCA	2220
Qy	2161	GATTTGAGGCTACACAGTATCTTTAAACGGGAAGGGTAAAGTCTGGAATPATCATTTCA	2260
Db	2221	TCAGTGAATTAATTTAACCCACAATCTTAGTGGCACAATTTAACATATCTCGGGAATATAACA	2280
Qy	2221	TCAGTGAATTAATTTAACCCACAATCTTAGTGGCACAATTTAACATATCTCGGGAATATAACA	2340
Db	2281	ATTTAACCAACTACGAGANAAGAACCTCGTATTGGCAACACAGCCATGATTGCGACATGG	2340
Qy	2281	ATTTAACCAACTACGAGANAAGAACCTCGTATTGGCAACACAGCCATGATTGCGACATGG	2400
Db	2341	ACGTCAGTGTCTTAATCTAGACAGAGGCGCAATTTTACCTTTTATTAATACATTTTCA	2400
Qy	2341	ACGTCAGTGTCTTAATCTAGACAGAGGCGCAATTTTACCTTTTATTAATACATTTTCA	2460
Db	2401	AGCAATAGCAAGGCTTAACACACAGTATAGAAGCTCTCAGGGGTGAATTTTAAACGGC	2460
Qy	2401	AGCAATAGCAAGGCTTAACACACAGTATAGAAGCTCTCAGGGGTGAATTTTAAACGGC	2520
Db	2461	GTAATGGCAACATGTCATTCAATCTCAAGAGGAGCGAAAGTTAATTTCAAAATTAATAA	2520
Qy	2461	GTAATGGCAACATGTCATTCAATCTCAAGAGGAGCGAAAGTTAATTTCAAAATTAATAA	2580
Db	2521	CCAAACGAGAACATGAACACAAAGCAACCTTTACCAATTCGGTTTTTACGCCAATATCACA	2580
Qy	2521	CCAAACGAGAACATGAACACAAAGCAACCTTTACCAATTCGGTTTTTACGCCAATATCACA	3020
Db	2581	GCACCTGTGGGGCTCTGTTTTTTTTTGATATATATGCCCAACCATCTTGGCAGAGGGCT	2640
Qy	2581	GCACCTGTGGGGCTCTGTTTTTTTTTGATATATATGCCCAACCATCTTGGCAGAGGGCT	2680
Db	2641	GAGTTTAAAAATTCAGTGAATTAATCTCTAACGGCGCTAAATTTTACCTTAAATTCCTCAT	2700
Qy	2641	GAGTTTAAAAATTCAGTGAATTAATCTCTAACGGCGCTAAATTTTACCTTAAATTCCTCAT	2740
Db	2701	GTTTCGGCGCGATGACGCTTTTAAAAATCAACAAGACTTAAACCAATAATGCAACCAATTTCA	2760
Qy	2701	GTTTCGGCGCGATGACGCTTTTAAAAATCAACAAGACTTAAACCAATAATGCAACCAATTTCA	2820
Db	2761	AATTTTCAGCCTCAGACAGCAAGAGATGATTTTTATGACGGGTACGACGCAATGCGCATC	2820
Qy	2761	AATTTTCAGCCTCAGACAGCAAGAGATGATTTTTATGACGGGTACGACGCAATGCGCATC	2880
Db	2821	AATTTCAACCTTACAACATATCCATCTTGGCGGTAAATGTCACCTTGGTGACAAAACTCA	2880
Qy	2821	AATTTCAACCTTACAACATATCCATCTTGGCGGTAAATGTCACCTTGGTGACAAAACTCA	2940
Db	2881	AGCAGCAGCATTAACGGGGAATATTACTATCAGAGAAGCAGCAAAATGTTACGCTAGAAGCC	2940
Qy	2881	AGCAGCAGCATTAACGGGGAATATTACTATCAGAGAAGCAGCAAAATGTTACGCTAGAAGCC	3000
Db	2941	AATAACGCCCTTAATCAGCAAAACATAGGGAATAGAGTTATAAACTTTGGCAGCTTGCTC	3000
Qy	2941	AATAACGCCCTTAATCAGCAAAACATAGGGAATAGAGTTATAAACTTTGGCAGCTTGCTC	3060
Db	3001	GTTAATGGGAGTTTAAGTTTAACCTGGGGAATGAGATATTAAGGCAATCTCACTATT	3060
Qy	3001	GTTAATGGGAGTTTAAGTTTAACCTGGGGAATGAGATATTAAGGCAATCTCACTATT	3120
Db	3061	TCAGAAAACGCCACTTTTAAAGGAAGACTTAGAGATACCTTAATATCACCGCAATTTT	3120
Qy	3061	TCAGAAAACGCCACTTTTAAAGGAAGACTTAGAGATACCTTAATATCACCGCAATTTT	3180
Db	3121	ACCAATTAATGGCAGCTGCGGAATTAATTAACCAAGGAGTGTGTAACCTTGGCAATGTT	3180
Qy	3121	ACCAATTAATGGCAGCTGCGGAATTAATTAACCAAGGAGTGTGTAACCTTGGCAATGTT	3240
Db	3181	ACCAATGATGGTGATTTAAACATTAACCTACCTCAGCTTAACGCAACCAAGAACATCATC	3240
Qy	3181	ACCAATGATGGTGATTTAAACATTAACCTACCTCAGCTTAACGCAACCAAGAACATCATC	3300
Db	3241	GGCGGAGATATATCAACAAAAAGGAGCTTAATATATACAGACAGTAAATATGATGCT	3300
Qy	3241	GGCGGAGATATATCAACAAAAAGGAGCTTAATATATACAGACAGTAAATATGATGCT	3360
Db	3301	GAATTCCAAATTTGGCGGCAATATCTCGCAAAAAGCAACCTCAGGATTTCTTCCGAT	3360
Qy	3301	GAATTCCAAATTTGGCGGCAATATCTCGCAAAAAGCAACCTCAGGATTTCTTCCGAT	3420
Db	3361	AAAAATTAATATCACCAACAGATTAACAAATCAAAAAGGTTATGATGGAGAGGACTTAGT	3420
Qy	3361	AAAAATTAATATCACCAACAGATTAACAAATCAAAAAGGTTATGATGGAGAGGACTTAGT	3480
Db	3421	TCAGATGCGACAAG	

QY 3661 AAAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCGACACGATACCGGCTTAAC 3720
 Db 4101 AAAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCGACACGATACCGGCTTAAC 4160
 QY 3721 ATTACTGCAAAAATGTAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATAATC 3780
 Db 4161 ATTACTGCAAAAATGTAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATAATC 4220
 QY 3781 ACCGGCTGGAAGGTTACCAACAGCAGCGCTCGACCAATTAACGCAACAATGGCAAA 3840
 Db 4221 ACCGGCTGGAAGGTTACCAACAGCAGCGCTCGACCAATTAACGCAACAATGGCAAA 4280
 QY 3841 GCAAGTATTACCAACCAACAGGTATACGCGGTACGATTTCCGGTAACACGTAAGT 3900
 Db 4281 GCAAGTATTACCAACCAACAGGTATACGCGGTACGATTTCCGGTAACACGTAAGT 4340
 QY 3901 GTTAGCGGACTGGTGATTTAAACCACTAAATCCGCTCAAAATTTGAAGCAAAATCGG 3960
 Db 4341 GTTAGCGGACTGGTGATTTAAACCACTAAATCCGCTCAAAATTTGAAGCAAAATCGG 4400
 QY 3961 GAGGCTAAATGTAACAGTGCACAGGTACAATTCGCGGTACAATTTCCGGTAATACGGTA 4020
 Db 4401 GAGGCTAAATGTAACAGTGCACAGGTACAATTTCCGGTAATACGGTAATACGGTA 4080
 QY 4021 AATGTTAGCGCAACCGCTGGCGATTTAACAGTTGGGATGGCGAGCAAAATTAATGCGACA 4080
 Db 4461 AATGTTAGCGCAACCGCTGGCGATTTAACAGTTGGGATGGCGAGCAAAATTAATGCGACA 4140
 QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTTGACTACTGAAGCGGTTCTAGC 4140
 Db 4521 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTTGACTACTGAAGCGGTTCTAGC 4580
 QY 4141 ATCACTTCACTAAGGTCAGGTAGACTCTTGCGCTCAGAATGGTACGATCCAGGAAGC 4200
 Db 4581 ATCACTTCACTAAGGTCAGGTAGACTCTTGCGCTCAGAATGGTACGATCCAGGAAGC 4260
 QY 4201 ATTAATGCTGTAATGTGACATTAATACCTACAGGCACCTTAACCAACCGTGGCAGGCTCG 4260
 Db 4641 ATTAATGCTGTAATGTGACATTAATACCTACAGGCACCTTAACCAACCGTGGCAGGCTCG 4320
 QY 4261 GATATTAACCAACAGCGGCACTTGGTATTAAAGCAAAAGATGCTAAAGCTAAATGGT 4320
 Db 4701 GATATTAACCAACAGCGGCACTTGGTATTAAAGCAAAAGATGCTAAAGCTAAATGGT 4380
 QY 4321 GATGCATCAGGTGATAGTACAGAGTGAATGCAAGTCAACGCAAGCGCTCGTAGTGTG 4380
 Db 4761 GATGCATCAGGTGATAGTACAGAGTGAATGCAAGTCAACGCAAGCGCTCGTAGTGTG 4440
 QY 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGCGGATTTAAACACAGTAATAGCGTTA 4440
 Db 4821 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGCGGATTTAAACACAGTAATAGCGTTA 4480
 QY 4441 AATATCATTTGAAAGATGGTGAAGCAACTGCGCTTAAGAGGCAAGGAATTTGAGGTG 4500
 Db 4881 AATATCATTTGAAAGATGGTGAAGCAACTGCGCTTAAGAGGCAAGGAATTTGAGGTG 4560
 QY 4501 AATATATCCAGCGAGGTGTAGCAAGTGTAGAGGAATTAATGAAGCAAGCAAGCGCTT 4560
 Db 4941 AATATATCCAGCGAGGTGTAGCAAGTGTAGAGGAATTAATGAAGCAAGCAAGCGCTT 5000
 QY 4561 GAAAAGTAAAGATTTATCTGATGAAGAAAGAAAGAACATTAAGCTTTGGTGTAACT 4620
 Db 5001 GAAAAGTAAAGATTTATCTGATGAAGAAAGAAAGAACATTAAGCTTTGGTGTAACT 4680
 QY 4621 GCTGTAGCTTTGTTAGCCAAATAATACAAATTAACAGTCAATACACAAATGAATTTACA 5120
 Db 5061 GCTGTAGCTTTGTTAGCCAAATAATACAAATTAACAGTCAATACACAAATGAATTTACA 5180
 QY 4681 ACCAGCCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 4740
 Db 5121 ACCAGCCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 5180
 QY 4741 GCGCAGGAGTATGTACCAATGTTGCTGACGATGACAGCGCTAGTCAAGTAATTCACAAG 4800

Db 5181 GCGCAGGAGTATGTACCAATGTGTGAGAGTGGACCGGTAGTCAATTTGACAAG 5240
 QY 4801 GTAGATTTTCATCTGCAATGAAGTCAATTTATTTTCTGATTTACTGTGGGTAA 4860
 Db 5241 GTAGATTTTCATCTGCAATGAAGTCAATTTATTTTCTGATTTACTGTGGGTAA 5300
 QY 4861 GTTCAGTACGGGCTTTACCCATCTTGTAAATAATACGAGAAATACAATAAGATTTTT 4920
 Db 5301 GTTCAGTACGGGCTTTACCCATCTTGTAAATAATACGAGAAATACAATAAGATTTTT 5360
 QY 4921 AACAGGTTATTATTATG 4937
 Db 5361 AACAGGTTATTATTATG 5377

RESULT 14
 HIU08876
 LOCUS
 DEFINITION
 Haemophilus influenzae 12 adhesin (hmw1A) gene and putative
 accessory processing proteins (hmw1B) and (hmw1C) genes, complete
 cds.
 U08876 M84616
 U08876.1 GI:475770
 SOURCE
 Haemophilus influenzae.
 Haemophilus influenzae
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.
 1 (bases 1 to 5116)
 Barenkamp, S.J. and Leininger, E.
 Cloning, expression, and DNA sequence analysis of genes encoding
 nonpeptide Haemophilus influenzae high-molecular-weight
 surface-exposed proteins related to filamentous hemagglutinin of
 Bordetella pertussis
 Infect. Immun. 60 (4), 1302-1313 (1992)
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 1548058
 2 (bases 5117 to 9221)
 Barenkamp, S.J. and St Geme, J.W. III.
 Genes encoding high-molecular-weight adhesion proteins of
 nonpeptide Haemophilus influenzae are part of gene clusters
 Infect. Immun. 62 (8), 3320-3328 (1994)
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 8039903
 3 (bases 1 to 9221)
 Barenkamp, S.J.
 Direct Submission
 Submitted (14-APR-1994) Stephen J. Barenkamp, Department of
 Pediatrics, St. Louis University School of Medicine, Cardinal
 Glennon Children's Hospital, 1465 South Grand Boulevard, St. Louis,
 MO 63104-1095, USA
 COMMENT
 On Apr 29, 1994 this sequence version replaced gi:148902.
 FEATURES
 Location/Qualifiers
 1..9221
 /organism="Haemophilus influenzae"
 /strain="12"
 /db_xref="taxon:727"
 /clone_lib="Lambda EMBL3 genomic library of strain 12"
 129..240
 /note="direct"
 /citation={1}
 /rpt_type=tandem
 351..4961
 /gene="hmw1A"
 351..4961
 /gene="hmw1A"
 /note="surface-associated protein"
 /citation={1}
 /codon_start=1
 /transl_table=11
 /product="adhesin"
 /protein_id="AAA20527.1"

repeat_region

gene

CDS

Db	2074	AAGTATCGCCTTTTGAGAAAGGACCAAGTC-----ATTACAGGTC	2118
Qy	2129	AGGCACTGTAACCATTTACAGGAGGGAAGATTTTCAGGCGCTAACACGATCTTTTAA	2188
Db	2119	AAGGGACTATTACC---TCAGGCAATCAAAAAGGTTTATAGATTTAATATGTCCTCTAA	2175
Qy	2189	ACGGAACGGGTAAAGCTCTGAATATCATTTTCATCACTAGTGAATA-----TTTAAACC	2239
Db	2176	ACGGCACTGGCAGCGACTGCAATTCCACACTTAAAGAACCAATAATACGTAACAA	2235
Qy	2240	ACAATCTTATGTGGCACAATTAACATATCTGGGAATATAACAAT---TAACCAAACTACGA	2296
Db	2236	ATAAATTTGAAGGACTTTAAATATTTCAAGGAGCGACTTACTGGAATTTAACCTCCCTAA	2295
Qy	2297	GAAGAACACACCTCGTATTGGCAACACGCCATGATTCGCACTGGAACGTCAGTCTCTTA	2356
Db	2296	AAATGAAGTGGATATGATAAATTCAAAGGAGCGACTTACTGGAATTTAACCTCCCTAA	2355
Qy	2357	ATCT---AGACAGAGGGCGAAATTTTACCTTTATTAAATACATTTCAAGCAATAGCAAA	2413
Db	2356	ATGTTCCGAGAGTGGCGAGTTTAACTCTCACTATTGACTCCAGGAGGAAGCGATAGTCAG	2415
Qy	2414	GCTTAAACAACACGATATAGAGCTCTGACGGGTGAATTTAACGGCGTAAATGGCAACA	2473
Db	2416	GCACACTTACCAGCCTTATATTTTAAACGGTATATCATTTCAACAAGACACT---ACCT	2472
Qy	2474	TGTCATTCATCTCAAGAGGAGCGGAAAGTTTAATTTCAAAATTAACCAACACGAGAACA	2533
Db	2473	TTAATCTTGAGCGAAATGCAAGAGTCAACTTTTGACATCAAGCGACCAATAGGATTAATA	2532
Qy	2534	TGAACAACAAGCAAACTTTTACCAATTCGGTTTTAGCCATATACAGCCACTGGTGGGG	2593
Db	2533	AGTATCTAGTTTGAATTACGCATCATTTATGGAACACTTTCAGTTTCGGGAGGGGA	2592
Qy	2594	GCTCTGTTTT-----TTTTTGATATATGCAACCACTTCCTGGCAGAGGGCTCAGTTAA	2647
Db	2593	GTGTTGATTTCCACACTTCTCGCTCATCTCTTAACGTCCTAAACGTCCTAAACGTCCT	2652
Qy	2648	AAATGAGTGAATTAATATCTCTTAACGGCGTAAATTTTACCTTAAATCCCATGTTGCGG	2707
Db	2653	ATTCTAAATCTTTTATGTTTCAACAGGGTCAAGTTTAAAGTTTAAACCTCAGGCTCAA	2712
Qy	2708	CGATGAGCTTTTAAATCAACAAGACATTAACCAATAATGCAACCAATTCAAATTTCA	2767
Db	2713	CAAAACCTGGCTCTCAATAGAGAAAGATTTACTTTAATGGCCAC---GGAGGCA	2766
Qy	2768	GCCTCAGACAGACGAAGATGATTTTATGACGGGTACGACGCAATGCAATTCATCAAA	2827
Db	2767	ACATAACACTTTTGCAGTTGAAGGCCGATGGAATGATTTGTTAAAGGCAATTTAGCCA	2826
Qy	2828	CCTACAACTATCCATTCCTGGCGGCTAATGTCAACCTTTGGTGGCAACAACATCAAGCA	2887
Db	2827	AAAAAACAATAACCTTTGAAGGAGGTAAATCAACCTTTGGCTCAGGAAGCCGTAAACG	2886
Qy	2888	GCATTACGGGGAATTTACTATCGAGAAGCAGCAAAATGTTACGCTAGAAGCCCAATACG	2947
Db	2887	AAATCGAAGGCAATGTTACTATCAACACGCTAAGCTCACTTTATCGTTTCGGATT	2946
Qy	2948	CCCTTAATCAGCAACAATAGGATAGAGTTATAAACTTGGCAGCTTGCCTGTTAATG	3007
Db	2947	TTGACACCATCAAA-----ACCTTTACTATTAAAAAAGATGTCATCAATTAATAGCG	3000
Qy	3008	GGAGTTTAAGTTTAACTGGGGAATCCAGATATAAGGCAATCTCACTATTTCAGAA	3067
Db	3001	GCAACCTTACCGCTGGAGGCAATATTGTCATATACCGGAAATCTTACCGTTGAAAGTA	3060
Qy	3068	GCGCCACTTTTAAAGGAAGACATAGAGATACCCATAATATCAACCGCAATTTTACCAATA	3127
Db	3061	ACGCTAATTTCAAAGCTATCACAAATTTTCACTTTTAAATGAGCGGCTGTTTGACAACA	3120
Qy	3128	ATGGCACTGCCGAATTAATATAACAACAAGAGTGGTAAACCTTGGCAATGTTACCAATG	3187
Db	3121	AAGCAATTTCAATATTTCCATTTGCCAAAGGAGGGCTCGCTTTAAAGACATTTAATAAT	3180

QY	3188	ATGGTGATTAAACATTACCACTCACGCTAAAGCGCAACCAAGAGCATCATCGCGGAG	3247
Db	3181		3240
QY	3248	ATATATCAACAAAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCTGAAATCC	3307
Db	3241		3300
QY	3308	AAATTGGCGGAATATCTCGCAAAAGAGGCAACCTCACGATTCTCTCCGATAAAATTA	3367
Db	3301		3360
QY	3368	ATATCAACAAACAGATACAATCAAAAGGATATTGATGGAGAGGACTAGTTCAGATG	3427
Db	3361		3420
QY	3428	CGACAAGTAATGCCAACCTACTATTAAACCAAGAATTGAAATTTACAGAGAGCCTAA	3487
Db	3421		3480
QY	3488	GTATTTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTAACTATTG	3547
Db	3481		3540
QY	3548	GCAACAGTAATGACGGTAACAGCGGTGCGAAGCGCAAAACAGTAACCTTTTAAACAATGTTA	3607
Db	3541		3597
QY	3608	AGATTCAAAAATCTCTGCTGACGCTCACAATGTGACACTAAATAGCAAAAGTGAACAACT	3667
Db	3598		3657
QY	3668	CTAGCAGCAATGCGGACGCTGAAAGCAATAGCGACAGGATACCGGCTTAACTATTACTG	3727
Db	3658		3717
QY	3728	CANAANATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAAATATCACCGGT	3787
Db	3718		3777
QY	3788	C---GGAAAAGGTTACCAACAGCAGGCTCGACCATTAACGCAACAATGGCAA-----	3839
Db	3778		3837
QY	3840	-----	3839
Db	3838	AGATAACCGCTCAACACAGGTAGTATCCTAGTGGAAATTGATCCAGCTCTGGCTCTGTAA	3897
QY	3840	-----	3839
Db	3898	CACCTTACTGCAACGAGGCGCTCTTGTGTGAAGCAATATTTCGGGCAACACGTTACTG	3957
QY	3840	-----	3839
Db	3958	TTACTGCAATAGCGGTGCATTAAACCACTTTTGGCAGGCTTACAAATTAAGGAACCGAGA	4017
QY	3840	---AGCAAGTATTACAACCAAAACAGGTGATACAGCGGTACGATTTCCGGTAAACACGTAA	3898
Db	4018		4077
QY	3899	GTGTTAGCGGCACGTGTTTAAACCACTAAATTCGGGCTCAAAAATTTGAAGCGAAATCGG	3958
Db	4078		4137
QY	3959	GTGAGGCTTAATGTACAAAGTGCACAGGTACAAATTGGCGGTACAATTTCCGGTAAATACGG	4018
Db	4138	GCCAGGCTAACGTAACAAGTGCACAGGTACAAATTGTTGTTGACGATTTTCGGTAAATACGG	4197
QY	4019	TAATGTTACGGCAACGCTGGCGGNTTACAGTTGGGATCGCCAGCAAAATTAATCGGA	4078
Db	4198	TAATGTTACGGCAACGCTGGCGGNTTACAGTTGGGATCGCCAGCAAAATTAATCGGA	4257

Qy	4079	CAGAAGGAGCTGCAACCTTAA	CCGCAACAGGGAATACCTT	GACTACTTGAACCGCTTCTA	4138	
Db	4258	CAGAAGGAGCTGCAACCTTAA	CTACATCTCGGCAATTA	CTCCGAGCTAGTTTCA	4317	
Qy	4139	GCATCACTTCAACTAAGGGT	CAGGTAGACCTCTTGGCT	CAGAAATGGTAGCATCCAGGA	4198	
Db	4318	ACATTACTTCAGCCAAAGGT	CAGGTAAATCTTCAGCT	CAGGATGGTAGCGTTCGAGGA	4377	
Qy	4199	GCATTAAATGCTGCTAAAT	GTGACATTAATACTACAGC	ACCTTTAACCAACCGTGGCAGGCT	4258	
Db	4378	GTATTAATGCCCGCAAT	GTGACATTAATACTACAGG	CACCTTTAACTACCGCTGAAGGGTT	4437	
Qy	4259	CGGATTTAAAGCAACAGCG	CGCACCTTTGGTTATTAACG	CAAAAAGATGCTAAAGCTAAATG	4318	
Db	4438	CACAATTAAATGCAACAG	CGGTACCTTGGTTATTAACG	CAAAAAGACGCTGAGCTAAATG	4497	
Qy	4319	GTGATGCATCAGGTGATAGT	CACAAAGTGAATGCAGT	CAACGCAACGGCTCTGTGAGTG	4378	
Db	4498	GCAGCAGCATTGGTAA	CCACACAGCTGGTAAATG	CAACCAACGCAAAATGGCTCCGCGACGC	4557	
Qy	4379	TCAGCTCGCGCACCTCAAG	CAGCTGTGAATATCATCTG	GGGATTTAAACACACAGTAAATGGT	4438	
Db	4558	TAATCGGACACACCTCAAG	CAGAGTCAATCATCTGGG	ATTTAATCACAAATAAATGGAT	4617	
Qy	4439	TAATATCATTTTCGAAAG	ATGGTAGAACACACTGTG	CGCTTAGAGCGCAAGAAATTGAGG	4498	
Db	4618	TAATATCATTTTCAAAAA	ACGGTATTAACACCGCTA	CTGTTAAAGCGCTTAAATTTGATG	4677	
Qy	4499	TGAATATATTCACGCC	AGGTAGCAAGTGTAGA	AGAGTAATTTGAACGGCAACCGTCC	4558	
Db	4678	TGAATATACATTCAC	CGGGTATAGCAACGGT	AGATGAAGTAATTTGAACGGCAACCGCATCC	4737	
Qy	4559	TTGAAAAAGTAAAGAT	TTATCTGATGAAGAAG	AGAAACATTTAGCTATAA	CTTGGTGTTAA	4618
Db	4738	TTGAGAAGGTAAAGAT	TTATCTGATGAAGAAG	AGAGCGTTAGCTATAA	CTTGGCGTAA	4797
Qy	4619	GTGCTGTACGTTTTGT	TGAGCCAAATTAATACA	ATTACAGTCAATACACAAAATGAATTTA	4678	
Db	4798	GTGCTGTACGTTTTT	TGAGCCAAATAATACA	ATTACAGTCAATACACAAAATGAATTTG	4857	
Qy	4679	CAACACAGCGTCAAGT	GTATATTTCTGAAGT	TAAGGCTGCTTTCTCAAGTGGTA	4738	
Db	4858	CAACACAGCCATTAAG	TCGCAATAGTCATTTCT	GAAGCGAGGCGCTGTTTCTCAACACAGTG	4917	
Qy	4739	ATGGCCACAGCATCTG	ATCCATTTGCTGACG	ATGCAGCGGTAGTCAGTAATTTGACA	4798	
Db	4918	ATGGCCGACCGGTG	CGGTTAATATCGCT	GATAAACGGCGGTAGTCAGTAATTTGACA	4977	
Qy	4799	AGGTAGATTTTCACTC	CTGCATGAAGTCANTT	ATTTTTCGTATATTTACTGTCGGGTTA	4858	
Db	4978	AGGTAGATTTTCACTC	CTGCATGAAGTCAT	TTTTATTTTCGTATATTTACTGTCGGGTTA	5037	
Qy	4859	AAGTTCAGTACGGG	TTTTACCATCTTGT	TAAAAAATACGGGAANTACAAATGAATTTT	4918	
Db	5038	AAGTTCAGTACGGG	TTTTACCATCTTGT	TAAAAAATACGGGAANTACAAATGAATTTT	5097	
Qy	4919	TTAACAGGTTATTAT	TATG 4937			
Db	5098	TTAACAGGTTATTAT	TATG 5116			

RESULT	15
AR032347	
LOCUS	AF032347
DEFINITION	Sequence 1 from patent US 5869065.
ACTION	AF032347
VERSION	AF032347
VESION	AF032347.1 GI:5947952
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 5116)
AUTHORS	Barenkamp,S.J. and St. Geme,J.William. III.

TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 5869065-A 1 09-FEB-1999;
FEATURES Location/Qualifiers
source 1. 5116

BASE COUNT 1766 a 1008 c 1037 g 1305 t
ORIGIN

Query Match 54.3%; Score 2680.2; DB 6; Length 5116;
Best Local Similarity 73.5%; Pred. No. 0;
Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps 16;

QY 41 ATGACAAACAAATATACAAACCTTTTGGAGTCTATATGCAATATTTAAAAAAT 100
|||||
Db 41 ATGACAAACAAATATACAAACCTTTTGGAGTCTATATGCAATATTTAAAAAAT 99
|||||
QY 101 AGTATAATCCGCCATATAAATGGTATATCTTTTCATCTTTCACTTTTCATC 160
|||||
Db 100 AGTATAATCCGCCATATAAATGGTATATCTTTTCATCTTTTCATCTTTTCATC 159
|||||
QY 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220
|||||
Db 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219
|||||
QY 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 280
|||||
Db 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 279
|||||
QY 281 GAATGAAGAGGAGCTGACGACGCAATATGAATTAATTAATTTCACTTAACCT 340
|||||
Db 280 GAATGAAGAGGAGCTGACGACGCAATATGAATTAATTTCACTTAACCT 339
|||||
QY 341 TAGGAGAAATATGAACAGATATATCGTCTCAAAATTCAGCAAGCGCTGATCTTGG 400
|||||
Db 340 TAGGAGAAATATGAACAGATATATCGTCTCAAAATTCAGCAAGCGCTGATCTTGG 399
|||||
QY 401 TTGCTGTCTGAATTTGGACGGGTTGTGACCAATTCACAGAAAGGCTTCGCTATG 460
|||||
Db 400 TTGCTGTCTGAATTTGGACGGGTTGTGACCAATTCACAGAAAGGCTTCGCTATG 459
|||||
QY 461 TTACTATCTTTAGGTGTAAACCACTTAGGTTAAAGCCACTTCCGCTATGTACTATCT 520
|||||
Db 460 CTGCTCGCATGAAGTGGCTCACTTAGGTTAAAGCCACTTCCGCTATGTACTATCT 519
|||||
QY 521 TAGGTGTAAACATCTATCCCAATCTGTTTATGCAAGCGGCTTACAGGAATGGATGAT 580
|||||
Db 520 TAGGTGTAAACATCTATCCCAATCTGTTTATGCAAGCGGCTTACAGGAATGGATGAT 579
|||||
QY 581 TACAGGCGACAGCCACTATGCAAGTAGATGGTAATAAACCATTTATCCGCAACAGTGTG 640
|||||
Db 580 TACAGGCGACAGCCACTATGCAAGTAGATGGTAATAAACCATTTATCCGCAACAGTGTG 639
|||||
QY 641 ACGCTATCATTAATTTGGAACAATTTAAACATCGACCAAAATGAATGGTGGCTTTTAC 700
|||||
Db 640 ACGTATCATTAATTTGGAACAATTTAAACATCGACCAAAATGAATGGTGGCTTTTAC 699
|||||
QY 701 AAGAAACACAACTCCGCGTATTCACCGTGTACATCTAACCAAAATCCCAATTA 760
|||||
Db 700 AAGAAACACAACTCCGCGTATTCACCGTGTACATCTAACCAAAATCCCAATTA 759
|||||
QY 761 AAGGATTTTATGATCTTAACGACAGTCTTTTATCAACCCAAATGGTATCACAATAG 820
|||||
Db 760 AAGGATTTTATGATCTTAACGACAGTCTTTTATCAACCCAAATGGTATCACAATAG 819
|||||
QY 821 GTAAGACGAATTTAATACACTTAATGGCTTTTACGGCTTCTACGGCTAGACATTTCTAACG 880
|||||
Db 820 GTAAGACGAATTTAATACACTTAATGGCTTTTACGGCTTCTACGGCTAGACATTTCTAACG 879
|||||
QY 881 AAACATCAAGGCGCTAATTTTACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAA 940
|||||
Db 880 AAACATCAAGGCGCTAATTTTACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAA 939
|||||
QY 941 TTGTGAATCACGGTTTAAATTAATCTGCGGTAAGACGGCAGTGAATCTTATTTGGTGCA 1000
|||||

Db 940 TTGTGAATCACGGTTTAAATTAATCTGCGTAAAGACGGCAGTAAATCTTATTTGGTGGCA 999
|||||
QY 1001 AAGTGAATAAACAGGGTGTGATTAGCGTAAATGGTGGAGCATTCTTTACTCGCAGGC 1060
|||||
Db 1000 AAGTGAATAAACAGGGTGTGATTAGCGTAAATGGTGGAGCATTCTTTACTCGCAGGC 1059
|||||
QY 1061 AAAAAATCACCATCAGCATATATAAACCCCAACCATTTACTTACAGCATTTGCCCGCCTG 1120
|||||
Db 1060 AAAAAATCACCATCAGCATATATAAACCCCAACCATTTACTTACAGCATTTGCCCGCCTG 1119
|||||
QY 1121 AAAATGAAGCGGTCAATCTGGCGGATATTTTGGCAAGGCGGTAAACATTAATGTCCGTG 1180
|||||
Db 1120 AAAATGAAGCGGTCAATCTGGCGGATATTTTGGCAAGGCGGTAAACATTAATGTCCGTG 1179
|||||
QY 1181 CTGCGACTATTTCGAACCAAGTAACTTTCTGCTGATTCTGTAAAGCAAGATATAAGCG 1240
|||||
Db 1180 CTGCGACTATTTCGAACCAAGTAACTTTCTGCTGATTCTGTAAAGCAAGATATAAGCG 1239
|||||
QY 1241 GCAATATTGTTCTTTCCGCCAAAGAGGTGAAGCGGAAATTTGGCGGTCTAATTTCCGCTC 1300
|||||
Db 1240 GCAATATTGTTCTTTCCGCCAAAGAGGTGAAGCGGAAATTTGGCGGTCTAATTTCCGCTC 1299
|||||
QY 1301 AAAATCAGCAAGCTAAAGCGGCAAGCTGTATGATTACAGCGGATAAAGTCAATTAATAA 1360
|||||
Db 1300 AAAATCAGCAAGCTAAAGCGGCAAGCTGTATGATTACAGCGGATAAAGTCAATTAATAA 1359
|||||
QY 1361 CAGCTGCAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1420
|||||
Db 1360 CAGCTGCAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1419
|||||
QY 1421 AGCGCGCAAGGTAAAGCGGCAATTTCAATTTAGCAAGAAACCTCTTTAGAAAGAGGCT 1480
|||||
Db 1420 AGCGCGCAAGGTAAAGCGGCAATTTCAATTTAGCAAGAAACCTCTTTAGAAAGAGGCT 1479
|||||
QY 1481 CAACCATCAATGTATCAGGCAAGAAAGAGCGGACCGCTATTGTGTGGCGGATATTG 1540
|||||
Db 1480 CAACCATCAATGTATCAGGCAAGAAAGAGCGGACCGCTATTGTGTGGCGGATATTG 1539
|||||
QY 1541 CGTTAATTGACGGCAATTTAAAGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTT 1600
|||||
Db 1540 CGTTAATTGACGGCAATTTAAAGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTT 1599
|||||
QY 1601 TTGTGAGACATCGGGCATTTATTTATCCATTGACAGCAATGCAATTTGTTAAACAAGAG 1660
|||||
Db 1600 TTGTGAGACATCGGGCATTTATTTATTCATCAAGACCAATGCAATTTGTGAGCCAAAG 1659
|||||
QY 1661 AGTGGTTGTAGACCTGTATGATTAACAATTTGAAGCCGAGACCCCTTTCGAATAATA 1720
|||||
Db 1660 AGTGGTTGTAGACCGGATAATGTATCTTATTAATGCAAGAAACAGCAGGACGCAATA 1719
|||||
QY 1721 CCGGTATAAATGATGAATTTCCCAACAGGCGGTCAGCAAGCAGCCCTAAAAAATA 1780
|||||
Db 1720 CTTTCAAGACCATGATACACGGGATCGGGAATAGTGCCAGCACCCCAACAAACAA 1779
|||||
QY 1781 GCGAACTAAAAACACAGCTTAACCAATATCTTCAATTTATCTGAAAAACGCTTGA 1840
|||||
Db 1780 AAGAA---AAGACAAATTAACAAACAACTCTTTGAGAGTATCTAAAAAGAGTACCT 1836
|||||
QY 1841 CAATGAATATAACGCAATCAAGAAACCTTACCGTTAATAGCTCAATCAACATTCGGAAGA 1900
|||||
Db 1837 TTGTTTAAACATCTCTATCAACCCATCTATGTCAATAGTCTCCATTAATTT---ATCCA 1893
|||||
QY 1901 ACTCCCATTTAATTTCTCCATAGTAAAGGTGAGCGTGGCGGAGGCTTCAGATGTATGAG 1960
|||||
Db 1894 ATGGCAGCTTAACCTTTGGAGTGAGGCTCGGAGCGGTGGCGGCTTGAGATTAACAACG 1953
|||||
QY 1961 ATATTAC-----TTCCTAAAGCGGGAATTTAACCATTATTTCTGGCGGATGGG 2008
|||||
Db 1954 ATATTACCGCGGTGATGATACAGAGGTGCAAACTTAACAAATTTACTCAGCGCGCTGGG 2013
|||||
QY 2009 TTGATGTTTATAAAAAATTTACGCTTGTATCAGGGTTTTTAAATATTACCGCGCTTCG 2068
|||||

Db 2014 TTGATGTTTCAATAAAATATCTCACTCGGGGGCGCAAGGTAACATAAAACATTTACAGCTAAAC 2073
 Qy 2069 TAGCTTTTGAAGGTGGAATTAACAAAGCAGCGAGCGGGCAAAATGCTAAAAATTCGCCCC 2128
 Db 2074 AAGATATCGCTTTTGAGAAAGGAAGCAACCAAGTC-----ATTACAGGTC 2118
 Qy 2129 AGGCACTGTACCACTTACAGAGAGGAGGAAAAGATTTTCAGGGCTTAACACGCTATCTTTAA 2188
 Db 2119 AAGGCACTATTAAC---TCAGGCAATCAAAAAGGTTTGTAGATTTAATAATGCTCTCTAA 2175
 Qy 2189 AGGCAACGGGTAAAGGCTGTGATATCAATTCATCAGTCAATAA-----TTTAAACCC 2239
 Db 2176 AGGCACTGGCAGCGAGCTGCAATTCACCACTAAAGAACCAATAAAATACGCTATCAAA 2235
 Qy 2240 ACAATCTTAGTGGCACAATTAACATATCTGGGAATATTAACAT---TAACCAAACTACGA 2296
 Db 2236 ATAAATTTGAAGGGACTTTAAATATTTTCAGGAAAGTGAACATCTCAATGGTTTACCTA 2295
 Qy 2297 GAAAGAACACCTCGTATTTGGCAACACAGCCATGATTCGCACTGGAAGCTAGTCTCTTA 2356
 Db 2296 AAAATGAAAGTGGATATGATAATTAACAGGAGCGACTTACTGGAATTTAACTCTCTTA 2355
 Qy 2357 ATCT---AGAGACAGGCGCAAAATTTTACCTTTTAAATATACATTTCAAGCAATACGAAG 2413
 Db 2356 ATGTTTCCGAGAGTGGCGAGTTTAACTCTCACTATTTGACTCCAGAGGAAGCGATAGTCAG 2415
 Qy 2414 GCTTAAACAACACAGTATAGAACTCTGCAAGGGTGAATTTTAAACGGCTTAAATGCAACA 2473
 Db 2416 GCACACTTACCGCCTTTATATTTTAAACGGTATATCTTCAACAAAGACACT---ACCT 2472
 Qy 2474 TGTCATCAATCTCAAGAGGAGCGAAAGTTAATTTCAAAATTAACCAACAAAGCAACA 2533
 Db 2473 TTAATGTTGACCAATATGCAAGAGTCACTTTGACATCAAGGCACCAATAGGATAAATA 2532
 Qy 2534 TGAACACAAGCAACCTTTACCAATTTTCGGTTTATAGCCATATCACAGCACTGTTGGG 2593
 Db 2533 AGTATCTTAGTTGAATAGCGATCATTAATGGAACATTTTCAGTTTCGGAGGGGGA 2592
 Qy 2594 GCTCTGTTT-----TTTTGATATATATGCAACCACTTCGCGAGAGGGCTGAGTTAA 2647
 Db 2593 GTGTTGATTTTACACTTCTCGCTCATPCTTAACGTCGAACCCCGGTGATGTTATAA 2652
 Qy 2648 AAATGAGTGAATTAATATCTTAACGGCGCTAATTTTACCTTTAAATTTCCCATGTTCCGG 2707
 Db 2653 ATTCTAAATACCTTTAATGTTTCAACAGGGTCAAGTTTAAAGATTTAAACCTTCAGGCTCA 2712
 Qy 2708 GCGATGACGCTTTTAAATCAACAAGACTTTAACCAATAATGCAACCAATTTCAAAATTTCA 2767
 Db 2713 CAAAACCTGGCTTCTCAATAGAGAAAGATTTTAACTTTTAAATGCCACC-----GGAGGCA 2766
 Qy 2768 GCCTCAGACAGCAAGAAAGATTTTATGAGGGTACGCAACGCAATGCCATCAATTTCAA 2827
 Db 2767 ACATAACACTTTTGAAGTGAAGGCACCGATGGAATGATTTGTTAAAGGCATTTGAGCCA 2826
 Qy 2828 CCTCACAATATCATCTCTGGCGGTAAATGTCACCTTTGGTGGACAAACCAATCAAGCAGA 2887
 Db 2827 AAAAACAATTAACCTTTGAAGAGGTACATCACTTTTGCTCCAGGAAGCCGTACAG 2886
 Qy 2888 GCATTACGGGGAATATTACTATCGAGAAAGCAGCAATGTTACGCTAGAGCCCAATTAAG 2947
 Db 2887 AAATCGAAGGCAATGTTTACTATCAATTAACAGCTACGTCATCTTATCGGTTCCGGATT 2946
 Qy 2948 CCCCTAATCAGCAAAACATAAGGATAGATTATAAACTTTGGCAGCTTGCTCGTTAATG 3007
 Db 2947 TTGCAACCAATCAAAA-----ACCTTTAACTTTTAAAAAAGATGTCATCAATTAAGCG 3000
 Qy 3008 GGAGTTTAACTTAACTGGGAAATGAGATATTAAGAGCAATCTCACTATTTTCAGAAA 3067
 Db 3001 GCAACCTTACCGCTGGAGCAATATTTGCAATATATAGCCGGAATCTTACCCTTTGAAAGTA 3060
 Qy 3068 GCGCCACTTTTAAAGGAAGACTAGATACCTTAAATATCACGGCAATTTTACCATA 3127
 Db 3061 ACGCTAATTTCAAGCTATCACAAAATTTCACTTTTAAATGAGGGCGCTTTTGGACAACA 3120

Qy 3128 ATGCACTCGGCAATTAATATTAACACACAGGAGTGGTAAACCTTGGCAATGTTTACCAATG 3187
 Db 3121 AAGCAATTCAAATATTTCCATTTGCCAAAGGAGGGCTCGCTTTAAAGACATTTGATAAT 3180
 Qy 3188 ATGCTGATTTAAACATTTACCCTCACGCTTAACGCAACCAAGCAATCATCGCGGAG 3247
 Db 3181 CCAAGAAATTTAAGCATATCCACCAATCCAGCTCCACCTTACCACATTTATTAAGCGCA 3240
 Qy 3248 ATATAATCAACAAAAAGGAGCTTAAATATACAGACATAATATATGATGCTGGAATCC 3307
 Db 3241 ATATAACCAATAAAACGGTGATTTAAATATACGAAGAGGTTAGTACTGGAATGC 3300
 Qy 3308 AAATTTGGCGCAATATCTCGCAAAAAGAGCAACCTCACGATTTCTTCGATAAATA 3367
 Db 3301 AAATTTGGCGCAATCTCTCGCAAAAAGAGTATCTCACGATTTCTTCTGACAAATCA 3360
 Qy 3368 ATATCACCAACAGATACAATCAAAAGGGTATTTGATGGAGAGGACTAGTTCAGATG 3427
 Db 3361 ATATTACCAACAGATAACAATCAAGCAGGTGTTGATGGGGAATTTCCGATTCAGACG 3420
 Qy 3428 CGAAGTAAATGCCAATCTAATTAACCAAAAGAAATGCAAAATTCACAGAACCTAA 3487
 Db 3421 CGCAACAATGCCAATCTAATTAACCAAAAGAAATGCAAAATTAACGCAAGACCTAA 3480
 Qy 3488 GTATTTGAGTTTCAATTAAGCAGAGATTAACGCAAAAGTGTAGAGATTTAATATG 3547
 Db 3481 ATATTTGAGTTTCAATTAAGCAGAGATTAACGCAAAAGTGTAGAGATTTAATATG 3540
 Qy 3548 GCAACAGTAAATGAGGTTAACAGCGGTGCGGAAGCAACCAAGTAACTTTTAACAATGTA 3607
 Db 3541 GTAACCAATATAGTGTCTGA---TGGTAAATGCAAAAGTAACTTTTAAACAGGTTA 3597
 Qy 3608 AAGATTTCAAAATCTCTCTGACGCTCAATATGTACACTAAATAGCAAAAGTGAACAT 3667
 Db 3598 AAGATTTCAAAATCTCTCTGACGCTCAACAGGTGACACTACACAGCAAGTGAACAT 3657
 Qy 3668 CTAGCAGCAATGGCGGAGTGAAGCAATAGCAGCAAGCAATACCGGCTTAACTATTA 3727
 Db 3658 CCGGTAGTAAATAACAACACTGAAGATAGCAGTGAACAATATGCGGCTTAACTATTA 3717
 Qy 3728 CAAAATGTAAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATCAACCGGT 3787
 Db 3718 CAAAATGTAAGTAAACAAAGATATTTACTTCTCTCAAAAGCAGTGAAGTCTCTGGA 3777
 Qy 3788 C---GGAAAGGTTACCAACAGCGCTCGACCAATTAACGCAACCAATGCA--- 3839
 Db 3778 CAAGTGAAGAAATTAACCAATAACAGGTACCAATTAACCGCAACCACTGTTAAGCTG 3837
 Qy 3840 ----- 3839
 Db 3838 AGATAACCGCTCAAAACAGTATGCTAGTGGAAATGAGTCCAGCTCTGCTCTGTA 3897
 Qy 3840 ----- 3839
 Db 3898 CACTTACTGCAACCGAGGCGCTCTTGCTGTGAAGCAATATTTTCGGCAACACCGTTACTG 3957
 Qy 3840 ----- 3839
 Db 3958 TTACTGCAATAGCGGTGCAATTAACCACTTTTGGCAGGCTCTACAATTAAGGAACCGAGA 4017
 Qy 3840 -AGCAAGTATTAACCAACCAAGGTGATATACGCGGTAGCTTTCGGTAAACAGGTA 3898
 Db 4018 GTGTAACTTCAAGTCAATCAGCGATATCGCGGTAGCTTCTGTTGGTGGCACAGTAG 4077
 Qy 3899 GTGTAGCGGCTGTTGATTTTAAACCACTAAATCCGGCTCAAAATTAAGCCCAATCGG 3958
 Db 4078 AGGTAAAGCAACCGAAGTTTAACTCACTCAATCCCAATTTCAAAATTAAGCAACACAG 4137
 Qy 3959 GTGAGGCTAATGTAAACAGGTGCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAACGG 4018
 Db 4138 GCGAGGCTAACGTAAACAGGTGCAACAGGTACAATTTGGTGGTACGATTTTCCGGTAAACGG 4197

Search completed: March 22, 2003, 00:41:58
Job time : 12519 secs

us-10-092-880-3.rng

Tue Mar 25 08:39:01 2003

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:43:31 ; Search time 909 Seconds
(without alignments)
12231.150 Million cell updates/sec

Title: US-10-092-880-3
Perfect score: 4937
Sequence: 1 taatatatacaagataataa.....ttaacagggtattattatg 4937

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4937	100.0	4937	18	Non-typeable Haemo
2	4935.4	100.0	4937	14	Sequence encoding
3	4924	99.7	4940	21	Haemophilus influe
4	4913	99.5	4937	15	Haemophilus high m
5	4828.2	97.8	9323	14	Gene cluster for h
6	4820.2	97.6	9323	14	Non-typeable Haemo
7	3263	66.1	3263	21	DNA encoding Haemo
8	2680.2	54.3	5116	14	Sequence encoding
9	2680.2	54.3	5116	15	Haemophilus high m

10	2680.2	54.3	5116	18	AAT90994	Non-typeable Haemo
11	2680.2	54.3	5116	21	AAAS2195	Haemophilus influe
12	2677	54.2	9220	14	AAQ49508	Gene cluster for h
13	2565.4	52.0	9171	18	AAT90996	Non-typeable Haemo
14	2245	48.03	4803	18	AAT90993	Non-typeable Haemo
15	2203.8	44.6	4702	15	AAQ49511	Partial sequence e
16	2202.2	44.6	4702	14	AAQ49511	Haemophilus high m
17	1974	40.0	4794	18	AAT90992	Non-typeable Haemo
18	1591.8	32.2	2874	21	AAAS2188	DNA encoding H. in
19	1591.8	32.2	2892	21	AAAS2188	Haemophilus influe
20	1500.8	30.4	3240	21	AAAS2185	Haemophilus influe
21	1497.6	30.3	2997	21	AAAS2191	Haemophilus influe
22	1484.4	30.1	3222	21	AAAS2186	DNA encoding H. in
23	1481.2	30.0	2979	21	AAAS2192	DNA encoding Haemo
24	1473.2	29.8	4287	15	AAQ72295	Partial sequence e
25	1469.2	29.8	4286	14	AAQ49510	Haemophilus influe
26	1414	28.6	2928	21	AAAS2177	Haemophilus influe
27	1397.6	28.3	2910	21	AAAS2178	DNA encoding H. in
28	1298.4	26.3	3018	21	AAAS2182	DNA encoding H. in
29	1298.4	26.3	3036	21	AAAS2181	Haemophilus influe
30	1188.4	24.1	3033	21	AAAS2189	Haemophilus influe
31	1172	23.7	3015	21	AAAS2190	DNA encoding H. in
32	1125.4	22.8	3443	21	AAAS2196	DNA encoding Haemo
33	1081.2	21.9	3306	21	AAAS2183	Haemophilus influe
34	1068	21.6	3288	21	AAAS2184	DNA encoding Haemo
35	753.6	15.3	3543	21	AAAS2194	Haemophilus influe
36	753.6	15.3	3568	21	AAAS2193	Haemophilus influe
37	643.2	13.0	3667	21	AAAS2180	DNA encoding H. in
38	643.2	13.0	3685	21	AAAS2179	Haemophilus influe
39	631.8	12.8	3681	21	AAAS2175	Haemophilus influe
40	628.8	12.7	3663	21	AAAS2176	DNA encoding H. in
41	65	1.3	105	21	AAAS2165	Plasmid DS-1200-3
42	64.2	1.3	2270	23	ABL1982	Drosophila melanog
43	62.8	1.3	1830	21	AAAT42063	Haemophilus influe
44	61	1.2	105	21	AAAS2164	Plasmid DS-1200-3
45	60.8	1.2	112190	22	AH44801	Human GPCR protein

ALIGNMENTS

RESULT 1
AAT90995
ID AAT90995 standard; DNA: 4937 BP.
XX
AC AAT90995;
XX
DT 14-APR-1998 (first entry)
XX
DE Non-typeable Haemophilus high mol.wt. surface protein hmw2 gene.
DE
KW Non-typeable Haemophilus; high molecular weight surface protein;
KW Hmw2; hmw2 gene; immunogen; vaccine; otitis media; ss.
XX
OS Haemophilus influenzae strain 12.
XX
FH Key Location/Qualifiers
FT RBS 341..345
FT /tag= a
FT repeat_unit 129..135
FT /tag= b
FT repeat_unit /rpt_type= TANDEM
FT /tag= c
FT repeat_unit /rpt_type= TANDEM
FT /tag= d
FT repeat_unit /rpt_type= TANDEM
FT /tag= e
FT repeat_unit /rpt_type= TANDEM
FT /tag= f

```

FT      repeat_unit      /rpt_type= TANDEM
FT      164..170          /*tag= g
FT      repeat_unit      /rpt_type= TANDEM
FT      171..177          /*tag= h
FT      repeat_unit      /rpt_type= TANDEM
FT      178..184          /*tag= i
FT      repeat_unit      /rpt_type= TANDEM
FT      185..191          /*tag= j
FT      repeat_unit      /rpt_type= TANDEM
FT      192..198          /*tag= k
FT      repeat_unit      /rpt_type= TANDEM
FT      199..205          /*tag= l
FT      repeat_unit      /rpt_type= TANDEM
FT      206..212          /*tag= m
FT      repeat_unit      /rpt_type= TANDEM
FT      213..219          /*tag= n
FT      repeat_unit      /rpt_type= TANDEM
FT      220..226          /*tag= o
FT      repeat_unit      /rpt_type= TANDEM
FT      227..233          /*tag= p
FT      repeat_unit      /rpt_type= TANDEM
FT      234..240          /*tag= q
FT      CDS               /rpt_type= TANDEM
FT      352..478          /*tag= r
FT      /transl_except= (pos:451..453, aa:Ser)
FT      /transl_except= (pos:454..456, aa:Glu)
FT      /transl_except= (pos:457..459, aa:Lys)
FT      /transl_except= (pos:460..462, aa:Pro)
FT      /transl_except= (pos:463..465, aa:Ala)
FT      /transl_except= (pos:466..468, aa:Arg)
FT      /transl_except= (pos:469..471, aa:Met)
FT      /transl_except= (pos:472..474, aa:Val)
FT      /transl_except= (pos:475..477, aa:Val)
FT      /transl_except= (pos:478..480, aa:Arg)
FT      /transl_except= (pos:1621..1623, aa:Asp)
FT      /transl_except= (pos:1627..1629, aa:Phe)
FT      /transl_except= (pos:1633..1635, aa:Lys)
FT      /transl_except= (pos:1636..1638, aa:Asp)
FT      /transl_except= (pos:1651..1653, aa:Asp)
FT      /transl_except= (pos:1654..1656, aa:Ala)
FT      /transl_except= (pos:1675..1677, aa:Phe)
FT      /transl_except= (pos:1681..1683, aa:Asn)
FT      /transl_except= (pos:1687..1689, aa:Ser)
FT      /transl_except= (pos:1693..1695, aa:Asn)
FT      /transl_except= (pos:1711..1713, aa:Phe)
PN      W09736914-AL.
XX
PD      09-OCT-1997.
XX
PF      01-APR-1997; 97WO-US04707.
XX
PR      01-APR-1996; 96US-0617697.
XX
PA      (BARE/) BARENKAMP S J.
XX
PI      Barenkamp SJ;
XX
DR      WPI; 1997-503038/46.
DR      P-PSDB; AAW30294.
XX

```

```

PT      High molecular weight proteins of non-typeable Haemophilus
PT      Influenzae - useful for vaccine production
XX
PS      Claim 7; Page 71-73; 183pp; English.
XX
CC      This nucleic acid comprises the hmw2 gene of non-typeable
CC      Haemophilus influenzae strain 12 that encodes high molecular
CC      weight surface protein HMW2 (see AAW30294). A phage genomic
CC      library of strain 12 was screened for clones expressing high
CC      mol.wt. proteins using a high titre antiserum against HMWs.
CC      Strongly reactive clones were subcloned into T7 expression
CC      plasmid; all expressed either 120 kDa HMW2 or 125 kDa HMW1 (see
CC      AAW30293). The expressed proteins are truncated, starting at
CC      residue 442 of both full-length HMW1 and HMW2 gene products.
CC      Downstream processing requires the products of additional
CC      genes (see AAT90996 and AAT90997). Nucleotide sequences
CC      (see AAT90992-93) encoding 2 HMW proteins of non-typeable H.
CC      influenzae strain 5 (see AAW30291-92) have also been identified.
CC      The HMW proteins, conjugates and peptides can be used in,
CC      vaccines, as immunogens for preparation of antibodies and as
CC      antigens for detection of these antibodies. The nucleic acid
CC      sequences can be used as to prepare recombinant proteins and as
CC      probes for detection of related genes.
XX
SQ      Sequence 4937 BP; 1729 A; 948 C; 1009 G; 1251 T; 0 other;

Query Match      100.0%; Score 4937; DB 18; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGTATGACAAACACAAATTACAA 60
Db      1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGTATGACAAACACAAATTACAA 60
Qy      61 CACCTTTTTCAGCTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCATATAA 120
Db      61 CACCTTTTTCAGCTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCATATAA 120
Qy      121 AATGGTATAATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 180
Db      121 AATGGTATAATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 180
Qy      181 CTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 240
Db      181 CTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 240
Qy      241 CACATGAATCATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db      241 CACATGAATCATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Qy      301 GAACGCAATGATAAGTAAATTTAATTTGTTCAACTAACCTTAGGAGAAAATATGAACAAG 360
Db      301 GAACGCAATGATAAGTAAATTTAATTTGTTCAACTAACCTTAGGAGAAAATATGAACAAG 360
Qy      361 ATATATCGTCTCAAAATTCAGCAACGCCCTGAATGCTTTGGTGGTGTGTCTGAATTTGGCA 420
Db      361 ATATATCGTCTCAAAATTCAGCAACGCCCTGAATGCTTTGGTGGTGTGTCTGAATTTGGCA 420
Qy      421 CGGGGTGTGACCAATTCACAGAAAAGGCTTCCCGCTATGTTACTATCTTTAGGTGTAC 480
Db      421 CGGGGTGTGACCAATTCACAGAAAAGGCTTCCCGCTATGTTACTATCTTTAGGTGTAC 480
Qy      481 CACTTAGCGTTAAAGCCACTTTCCCGCTATGTTACTATCTTTAGGTGTATTCATTC 540
Db      481 CACTTAGCGTTAAAGCCACTTTCCCGCTATGTTACTATCTTTAGGTGTATTCATTC 540
Qy      541 CAATCTGTTTTAGCAAGCGGCTTACAAAGGAATGGATGTAGTACACGCGCACGCCACTATG 600
Db      541 CAATCTGTTTTAGCAAGCGGCTTACAAAGGAATGGATGTAGTACACGCGCACGCCACTATG 600
Qy      601 CAAAGTAGATGTTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 660
Db      601 CAAAGTAGATGTTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 660

```

us-10-092-880-3.rng

Tue Mar 25 08:39:01 2003

QY	661	CAATTTAACTCGACCAAAATGAATGGTCAGTGTCTTAAAGAAACAACTCGGC	720	1741	CCAAAGCGACGGTGAAGCAAGCGACCTTAAATAAATAAGCAACTCAAAACACGCTA	1800
Db	661	CAATTTAACTCGACCAAAATGAATGGTCAGTGTCTTAAAGAAACAACTCGGC	720	1741	CCAAAGCGACGGTGAAGCAAGCGACCTTAAATAAATAAGCAACTCAAAACACGCTA	1800
QY	721	GTATTTCAACCGTGTAGATCTTAACCAATCCCAATTAAGAGGATTTTAGATCTAAC	780	1801	ACCAATCAACTATTTCAAAATTTATCTGAAACACGCTTGAACAAATGAATATAAGCGCATCA	1860
Db	721	GTATTTCAACCGTGTAGATCTTAACCAATCCCAATTAAGAGGATTTTAGATCTAAC	780	1801	ACCAATCAACTATTTCAAAATTTATCTGAAACACGCTTGAACAAATGAATATAAGCGCATCA	1860
QY	781	GGCAAGTCTTTTAAATCAACCAANTGGTATCAATAGTAAAGACGAATTTATTAAAC	840	1861	AGAAACCTTACGGTTAATAGCTCAATCAACATCGAAGCAACTCCCACTTTAAATCTCCAT	1920
Db	781	GGCAAGTCTTTTAAATCAACCAANTGGTATCAATAGTAAAGACGAATTTATTAAAC	840	1861	AGAAACCTTACGGTTAATAGCTCAATCAACATCGAAGCAACTCCCACTTTAAATCTCCAT	1920
QY	841	ACTAATGGCTTTACGGCTTCTACGTAGACATTTCTACGAAACATCAAGCGCGTAAT	900	1921	AGTAAAGTTCAGCGTGGGAGGCGGTTTCAGATTGATGAGATATTTACTTCTAAAGCGGGA	1980
Db	841	ACTAATGGCTTTACGGCTTCTACGTAGACATTTCTACGAAACATCAAGCGCGTAAT	900	1921	AGTAAAGTTCAGCGTGGGAGGCGGTTTCAGATTGATGAGATATTTACTTCTAAAGCGGGA	1980
QY	901	TTCACTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGAATTCACGGTTAAAT	960	1981	AATTTAACCATTTATTTCTGGCGGATGGGTGATGTTTCATATAAATATTTACGCTTGATCAG	2040
Db	901	TTCACTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGAATTCACGGTTAAAT	960	1981	AATTTAACCATTTATTTCTGGCGGATGGGTGATGTTTCATATAAATATTTACGCTTGATCAG	2040
QY	961	ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGAGGAGGTGG	1020	2041	GGTTTTTTTAAATATTTACCGCCGCTTCCGCTAGCTTTTGAAGGTGGAAATTAACAAAGCAGC	2100
Db	961	ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGAGGAGGTGG	1020	2041	GGTTTTTTTAAATATTTACCGCCGCTTCCGCTAGCTTTTGAAGGTGGAAATTAACAAAGCAGC	2100
QY	1021	ATTAGCGTAATGGTGGCAGCATTTCTTACTCGCAGGCGCAAAATCACCATCAGCGAT	1080	2101	GACGCGCAAAATGCTTAAATTTGTCGCCAGGCGACTGTAAACCATTTACAGGAGAGGGA	2160
Db	1021	ATTAGCGTAATGGTGGCAGCATTTCTTACTCGCAGGCGCAAAATCACCATCAGCGAT	1080	2101	GACGCGCAAAATGCTTAAATTTGTCGCCAGGCGACTGTAAACCATTTACAGGAGAGGGA	2160
QY	1081	ATAATAACCCCAACCATTTACTTACAGCATTTGCGCGCTGAAATGAAGCGTCAATCTG	1140	2161	GATTTACAGGCTTAAACACGCTATCTTTAAACGGCAACGGTAAAGGCTCTGAATATCATTTCA	2220
Db	1081	ATAATAACCCCAACCATTTACTTACAGCATTTGCGCGCTGAAATGAAGCGTCAATCTG	1140	2161	GATTTACAGGCTTAAACACGCTATCTTTAAACGGCAACGGTAAAGGCTCTGAATATCATTTCA	2220
QY	1141	GGCGATTTTTCGCAAGCGGTACATTAATGTCGCTGCTGCCACTATTCGAAACCAA	1200	2221	TCAGTGAATTAATTAACCCCAACATCTAGTGGCACAATTAACATATCTCGGGAATATAACA	2280
Db	1141	GGCGATTTTTCGCAAGCGGTACATTAATGTCGCTGCTGCCACTATTCGAAACCAA	1200	2221	TCAGTGAATTAATTAACCCCAACATCTAGTGGCACAATTAACATATCTCGGGAATATAACA	2280
QY	1201	GGTAACTTTCTGCTGATCTGTAAGCAAGATTAAGCGGCAATATTTCTTTCCGCG	1260	2281	ATTAACCAACTACGAGAAAGACACCTGATTTGGCAACACGAGCCATGATTCGCACTGG	2340
Db	1201	GGTAACTTTCTGCTGATCTGTAAGCAAGATTAAGCGGCAATATTTCTTTCCGCG	1260	2281	ATTAACCAACTACGAGAAAGACACCTGATTTGGCAACACGAGCCATGATTCGCACTGG	2340
QY	1261	AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCAAGC	1320	2341	AACGTCAGTCTCTTAATCTAGAGACAGGCGCAATTTTACCTTTATTAATACATTTCA	2400
Db	1261	AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCAAGC	1320	2341	AACGTCAGTCTCTTAATCTAGAGACAGGCGCAATTTTACCTTTATTAATACATTTCA	2400
QY	1321	GGCAAGCTGATGATTTACAGCGGATAAAGTTCACATTTAAACACAGGTGAGTTATCGACTT	1380	2401	AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTCGAGGGGTGAATTTTAAACGGC	2460
Db	1321	GGCAAGCTGATGATTTACAGCGGATAAAGTTCACATTTAAACACAGGTGAGTTATCGACTT	1380	2401	AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTCGAGGGGTGAATTTTAAACGGC	2460
QY	1381	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGGTAAAC	1440	2461	GTAAATGGCAACATGTCATTTCAATCTCAAGAGGAGGCAAGGTTAATTTCAATTTAA	2520
Db	1381	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGGTAAAC	1440	2461	GTAAATGGCAACATGTCATTTCAATCTCAAGAGGAGGCAAGGTTAATTTCAATTTAA	2520
QY	1441	GGCATCAATTTAGCAAGAAACCTTTTAAAGAAAGCTTCAACCATCAATGTATCAGGC	1500	2521	CCAAACGAGAACATCAACACAGCAACCTTTTACCAATTCGGTTTTTATAGCAATATCACA	2580
Db	1441	GGCATCAATTTAGCAAGAAACCTTTTAAAGAAAGCTTCAACCATCAATGTATCAGGC	1500	2521	CCAAACGAGAACATCAACACAGCAACCTTTTACCAATTCGGTTTTTATAGCAATATCACA	2580
QY	1501	AAAGAAAAGCGGCGCTATTTGTTGGGCGATATTTGGTTAAATGACGCGCAATATT	1560	2581	GCCACTGGTGGGCGCTCTGTTTTTTTGTATATATATGCAACCACTTCTGGCAGAGGGCT	2640
Db	1501	AAAGAAAAGCGGCGCTATTTGTTGGGCGATATTTGGTTAAATGACGCGCAATATT	1560	2581	GCCACTGGTGGGCGCTCTGTTTTTTTGTATATATATGCAACCACTTCTGGCAGAGGGCT	2640
QY	1561	AACGCTCAAGGTAGTGGTATTCGCTAAACCGGTGTTTGTGGAGACATTCGGGCGAT	1620	2641	GAGTTAAAATGAGTGGAAATTAATCTCTAACGGCGTAAATTTTACCTTTAAATTTCCCAT	2700
Db	1561	AACGCTCAAGGTAGTGGTATTCGCTAAACCGGTGTTTGTGGAGACATTCGGGCGAT	1620	2641	GAGTTAAAATGAGTGGAAATTAATCTCTAACGGCGTAAATTTTACCTTTAAATTTCCCAT	2700
QY	1621	TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGTGGTGTGACGCTGAT	1680	2701	GTTTCGGGGGATGACGCTTTTAAATCAACAGACTTAACCATTAATGCAACCAATTTCA	2760
Db	1621	TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGTGGTGTGACGCTGAT	1680	2701	GTTTCGGGGGATGACGCTTTTAAATCAACAGACTTAACCATTAATGCAACCAATTTCA	2760
QY	1681	GATGTAACAATTTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATTAATGATGAATTC	1740	2761	AATTTACGCTCAGACAGAGCAAGAGATGATTTTATGACGGGTACGAGCAATGCCATC	2820
Db	1681	GATGTAACAATTTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATTAATGATGAATTC	1740	2761	AATTTACGCTCAGACAGAGCAAGAGATGATTTTATGACGGGTACGAGCAATGCCATC	2820
QY	1741	GATGTAACAATTTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATTAATGATGAATTC	1800	2821	AATTTACCTTACACATATCCATTTCTGGGCGGTAAATGTCACCCCTTGGTGGCAACCACTCA	2880
Db	1741	GATGTAACAATTTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATTAATGATGAATTC	1800	2821	AATTTACCTTACACATATCCATTTCTGGGCGGTAAATGTCACCCCTTGGTGGCAACCACTCA	2880

Db 2821 AATCAACCTACACATATCCATTCTGGCGGTAAATGTCACCCCTGGTGGACAAACTCA 2880
QY 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940
Db 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940
QY 2941 AATAACGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACCCTTGGCAGCTTGCTC 3000
Db 2941 AATAACGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACCCTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGAGTTTAAGTTTAACCTGGCGAAATGCAGATATAAGGCAATCTCACTATT 3060
Db 3001 GTTAATGGAGTTTAAGTTTAACCTGGCGAAATGCAGATATAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCAACCGGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCAACCGGCAATTTT 3120
QY 3121 ACCAATAATGGCACTGCGGAAATTAATATACACAGGAGTGTAAACHTTGGCAATGTT 3180
Db 3121 ACCAATAATGGCACTGCGGAAATTAATATACACAGGAGTGTAAACHTTGGCAATGTT 3180
QY 3181 ACCAATGATGGTATTTAAACATTTACACCTACCGCTAAACGCAACCAAGAGCATCATC 3240
Db 3181 ACCAATGATGGTATTTAAACATTTACACCTACCGCTAAACGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAGAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
Db 3241 GCGGAGATATAATCAACAAAGAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAAAGGCAACCTCACGATTTCTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAAAGGCAACCTCACGATTTCTCCGAT 3360
QY 3361 AAAATTAATATACCAAAACAGATACAAATCAAAAGGGTATTGATGGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATATACCAAAACAGATACAAATCAAAAGGGTATTGATGGAGAGGACTCTAGT 3420
QY 3421 TCAGATGCCACAAGTAATGCCAACCTAACTATTAAACCAAGAAATGAAATTTGACAGAA 3480
Db 3421 TCAGATGCCACAAGTAATGCCAACCTAACTATTAAACCAAGAAATGAAATTTGACAGAA 3480
QY 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
Db 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAAGCCAAACAGTAACCTTTTAAAC 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAAGCCAAACAGTAACCTTTTAAAC 3600
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTCAGCGTACAAATAGGCAACGATACCGGCTTAAC 3660
Db 3601 AATGTTAAAGATTCAAAAATCTCTGCTCAGCGTACAAATAGGCAACGATACCGGCTTAAC 3660
QY 3661 AAAACATCTAGCAATAGGAGTAAACAGATATTACTTCTCTCAAAACAGTAATAATC 3720
Db 3661 AAAACATCTAGCAATAGGAGTAAACAGATATTACTTCTCTCAAAACAGTAATAATC 3720
QY 3721 ATTACTGCAAAAATGTAAGTAAACAGATATTACTTCTCTCAAAACAGTAATAATC 3780
Db 3721 ATTACTGCAAAAATGTAAGTAAACAGATATTACTTCTCTCAAAACAGTAATAATC 3780
QY 3781 ACCGCGTGGAAAGGTTACCAACAGAGGCTCGACCATTAACGCAACAAATGGCAAA 3840
Db 3781 ACCGCGTGGAAAGGTTACCAACAGAGGCTCGACCATTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACACCAACAGAGTATATCAGCGGTACGATTTCCGGTAACACGGTAAGT 3900
Db 3841 GCAAGTATTACACCAACAGAGTATATCAGCGGTACGATTTCCGGTAACACGGTAAGT 3900
QY 3901 GTTAGCGGACTGGTATTTAACTAACTAACTCGGCTCAAAATTTAAGCGAAATCGGGT 3960
Db 3901 GTTAGCGGACTGGTATTTAACTAACTAACTCGGCTCAAAATTTAAGCGAAATCGGGT 3960

Db 3901 GTTAGCGGACTGGTATTTAACTAACTAACTCGGCTCAAAATTTAAGCGAAATCGGGT 3960
QY 3961 GAGGCTAATGTAAACAGTGCACAGGTACAAATTTGGCGGTACAATTTCCGGTAATACGGTA 4020
Db 3961 GAGGCTAATGTAAACAGTGCACAGGTACAAATTTGGCGGTACAATTTCCGGTAATACGGTA 4020
QY 4021 AATGTTACGCCAACAGCTGGCGATTTAAACAGTTTGGGAATGGCGCAGAAATTAATGGGACA 4080
Db 4021 AATGTTACGCCAACAGCTGGCGATTTAAACAGTTTGGGAATGGCGCAGAAATTAATGGGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGCTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGCTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTTGGCTTCAGAAATGGTAGCATCGCAGGAAGC 4200
Db 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTTGGCTTCAGAAATGGTAGCATCGCAGGAAGC 4200
QY 4201 ATTAATGCTGCTAATGTGACATTTAACTACAGGCACTTAACCACTGGCAGGCTCG 4260
Db 4201 ATTAATGCTGCTAATGTGACATTTAACTACAGGCACTTAACCACTGGCAGGCTCG 4260
QY 4261 GATATTAAGCAACAGCGGCACTTTGGTATTAAACGCAAAAGATGCTAAGCTAATGGT 4320
Db 4261 GATATTAAGCAACAGCGGCACTTTGGTATTAAACGCAAAAGATGCTAAGCTAATGGT 4320
QY 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
QY 4381 ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATGGGTTA 4440
Db 4381 ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATGGGTTA 4440
QY 4441 AATATCACTTCGAAGATGGTAGAAGCACTGTGCGCTTAAAGAGGCAAGAAATGAGGTG 4500
Db 4441 AATATCACTTCGAAGATGGTAGAAGCACTGTGCGCTTAAAGAGGCAAGAAATGAGGTG 4500
QY 4501 AAATATATCCAGCGAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCTT 4560
Db 4501 AAATATATCCAGCGAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCTT 4560
QY 4561 GAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTAAGAGGCAAGAAATGAGGTG 4620
Db 4561 GAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTAAGAGTGTAGCTT 4620
QY 4621 GCTGTACGTTTCTGAGCCAAATATAACATTTACAGTCAATACAGAAATGAATTTTACA 4680
Db 4621 GCTGTACGTTTCTGAGCCAAATATAACATTTACAGTCAATACAGAAATGAATTTTACA 4680
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 4740
Db 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 4740
QY 4741 GGCGCACAGTATGTACCAATGTCTGACGATGGACAGCCGTAGTCAAGTAAATGACAAG 4800
Db 4741 GGCGCACAGTATGTACCAATGTCTGACGATGGACAGCCGTAGTCAAGTAAATGACAAG 4800
QY 4801 GTAGATTTCACTCCTGCAATGAAGTCAATTTATTTTCGTTATTTTACGTTGGGTTAAA 4860
Db 4801 GTAGATTTCACTCCTGCAATGAAGTCAATTTATTTTCGTTATTTTACGTTGGGTTAAA 4860
QY 4861 GTTACAGTACGGCTTTACCCATCTTGTAAAAAATTTACGGAGAATAAATGATTTT 4920
Db 4861 GTTACAGTACGGCTTTACCCATCTTGTAAAAAATTTACGGAGAATAAATGATTTT 4920
QY 4921 AACAGTTTATTATG 4937
Db 4921 AACAGTTTATTATG 4937

QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540
DB 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540
QY 541 CAATCTGTTTTAGCAAGCGGCTTACAGGAATGGATGTAGTACACGGCACACCCACTATG 600
DB 541 CAATCTGTTTTAGCAAGCGGCTTACAGGAATGGATGTAGTACACGGCACACCCACTATG 600
QY 601 CAAGTAGATGGTAATAAACCACTTATCCGCAACAGTGTGACGCTATCATTAATTTGNA 660
DB 601 CAAGTAGATGGTAATAAACCACTTATCCGCAACAGTGTGACGCTATCATTAATTTGNA 660
QY 661 CAATTTAATCATCGACCAAAATGAATGTGCAAGTGTGACGCTATCATTAATTTGNA 720
DB 661 CAATTTAATCATCGACCAAAATGAATGTGCAAGTGTGACGCTATCATTAATTTGNA 720
QY 721 GTATTTCAACCGTGTATACATCTAACCAATCTCCCAATCTCCCAATTTAGATTCTTAAC 780
DB 721 GTATTTCAACCGTGTATACATCTAACCAATCTCCCAATTTAGATTCTTAAC 780
QY 781 GGACAAGTCTTTTTAATCAACCCAAATGGTATCAATAGGTAAAGACGCAATTTAATTAAC 840
DB 781 GGACAAGTCTTTTTAATCAACCCAAATGGTATCAATAGGTAAAGACGCAATTTAATTAAC 840
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGCGCGCTAAT 900
DB 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGCGCGCTAAT 900
QY 901 TTCACCTTCGAGCAACCAAGATAAGCGCTCGCTGAAATGTGATCAGCGTTTAAAT 960
DB 901 TTCACCTTCGAGCAACCAAGATAAGCGCTCGCTGAAATGTGATCAGCGTTTAAAT 960
QY 961 ACTGTCGGTAAAGACGCGAGTGAATCTTATTTGGTGGCAAGTGAAGACGAGGCTG 1020
DB 961 ACTGTCGGTAAAGACGCGAGTGAATCTTATTTGGTGGCAAGTGAAGACGAGGCTG 1020
QY 1021 ATTAGCTAAATGGTGGCAGCATTTCTTTACTGCGAGGCAAAAAATCACCATCAGCGAT 1080
DB 1021 ATTAGCTAAATGGTGGCAGCATTTCTTTACTGCGAGGCAAAAAATCACCATCAGCGAT 1080
QY 1081 ATAAATAAACCCCAACCATTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140
DB 1081 ATAAATAAACCCCAACCATTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140
QY 1141 GCGCATATTTTGGCAAGGCGGTAACTATTAATGTCGCTGCTGCACTATTCGAAACCAA 1200
DB 1141 GCGCATATTTTGGCAAGGCGGTAACTATTAATGTCGCTGCTGCACTATTCGAAACCAA 1200
QY 1201 GGTAACTTTCTGCTGATTTCTGTAGCAAGATAAAGCGGCAATATTTGTTCTTCGCGC 1260
DB 1201 GGTAACTTTCTGCTGATTTCTGTAGCAAGATAAAGCGGCAATATTTGTTCTTCGCGC 1260
QY 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTAAATTTCCGCTCAAAATCAGCAAGCTTAAAGC 1320
DB 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTAAATTTCCGCTCAAAATCAGCAAGCTTAAAGC 1320
QY 1321 GCGAAGCTGATGATTACAGCGGATTAAGTCAATTAATAACAGGTGCGATTTCGACCTT 1380
DB 1321 GCGAAGCTGATGATTACAGCGGATTAAGTCAATTAATAACAGGTGCGATTTCGACCTT 1380
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGGCAAGGTAAAGC 1440
DB 1381 TCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGGCAAGGTAAAGC 1440
QY 1441 GGCATTTCAATTTAGCAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGC 1500
DB 1441 GGCATTTCAATTTAGCAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGC 1500
QY 1501 AAAGAAAAAGCGGCGCTATTTGTTGGGCGATATTTGCGTTAATTCACGCGCAATATT 1560
DB 1501 AAAGAAAAAGCGGCGCTATTTGTTGGGCGATATTTGCGTTAATTCACGCGCAATATT 1560
QY 1561 AACGCTCAAGGTAGTGGTATTCGCTAAACCGGTGGTTTTGTGGAGACATCGGGCAT 1620

DB 1561 AACGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTTATTCATTTGACAGCAATGCAATTTGTTAAACAAAAAGAGTGGTTGCTAGACCCTGAT 1680
DB 1621 TATTTATTCATTTGACAGCAATGCAATTTGTTAAACAAAAAGAGTGGTTGCTAGACCCTGAT 1680
QY 1681 GATGTAAACAATTTGAAGCGGAGAGCCCTTCGCAATAATACCCTGATTAATGATGATTC 1740
DB 1681 GATGTAAACAATTTGAAGCGGAGAGCCCTTCGCAATAATACCCTGATTAATGATGATTC 1740
QY 1741 CCAACAGGCAACCGGTGAAGCAAGCGCCCTTAAAAAAATAGCGAACTCAAAACAGCTA 1800
DB 1741 CCAACAGGCAACCGGTGAAGCAAGCGCCCTTAAAAAAATAGCGAACTCAAAACAGCTA 1800
QY 1801 ACCAATACAACTATTTCAAAATTTATCTGAAACAGCGCTGGACAATGAATATAACGCGATCA 1860
DB 1801 ACCAATACAACTATTTCAAAATTTATCTGAAACAGCGCTGGACAATGAATATAACGCGATCA 1860
QY 1861 AGAAACTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920
DB 1861 AGAAACTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920
QY 1921 AGTAAAGGTACGCGTGGCGGAGCGCTTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1980
DB 1921 AGTAAAGGTACGCGTGGCGGAGCGCTTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1980
QY 1981 AATTTAACCACTTATTTCTGGCGGTGGCTTGTATGTTTCAATAAAATATTACCTTTGATCAG 2040
DB 1981 AATTTAACCACTTATTTCTGGCGGTGGCTTGTATGTTTCAATAAAATATTACCTTTGATCAG 2040
QY 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATAACAAAGCAGC 2100
DB 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATAACAAAGCAGC 2100
QY 2101 GACGCGCAATGCTAAATTTGTCGCCAGGCACTGTACACCATTAACAGAGAGGAGAA 2160
DB 2101 GACGCGCAATGCTAAATTTGTCGCCAGGCACTGTACACCATTAACAGAGAGGAGAA 2160
QY 2161 GATTTACAGGCTAACACAGTATCTTTTAAAGGTGGAATAACAAAGCAGC 2220
DB 2161 GATTTACAGGCTAACACAGTATCTTTTAAAGGTGGAATAACAAAGCAGC 2220
QY 2221 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGCAATTAACATATCTGGGAATATAACA 2280
DB 2221 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGCAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAAACTACGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGG 2340
DB 2281 ATTAACCAAACTACGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGG 2340
QY 2341 AACGTCAGTCTCTTAATCTTAGAGCAGCGCAATTTTACCTTTTAAATTAATTTTAA 2400
DB 2341 AACGTCAGTCTCTTAATCTTAGAGCAGCGCAATTTTACCTTTTAAATTAATTTTAA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGAGGGGTGAATTTAACGCG 2460
DB 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGAGGGGTGAATTTAACGCG 2460
QY 2461 GTAAATGGCAACATGTCATTCAATCTCAAGAGGAGGGAAGTTAATTTCAATTTAA 2520
DB 2461 GTAAATGGCAACATGTCATTCAATCTCAAGAGGAGGGAAGTTAATTTCAATTTAA 2520
QY 2521 CCAACAGGAGACATGAACACAGCAAGCACTTTACCAATTCGCTTTTACCAATATACA 2580
DB 2521 CCAACAGGAGACATGAACACAGCAAGCACTTTACCAATTCGCTTTTACCAATATACA 2580
QY 2581 GCGACTGTGGGGCTCTGTTTTTGTATATATATGCAACCATTTCTGCGAGAGGGCT 2640
DB 2581 GCGACTGTGGGGCTCTGTTTTTGTATATATATGCAACCATTTCTGCGAGAGGGCT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGCGCTAATTTTACCTTAAATTTCCAT 2700
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGCGCTAATTTTACCTTAAATTTCCAT 2700

Db 2641 GAGTTAAAAAGTGAATAATATCTCTAACGGCGCTAATTTTACCTTAAATTCCTCAT 2700
Qy 2701 GTTCGCGCGATGAGCTTTTAAATAATCAACAAAGACTTTAAACCAATAATGCAACCAATTC 2760
Db 2701 GTTCGCGCGATGAGCTTTTAAATAATCAACAAAGACTTTAAACCAATAATGCAACCAATTC 2760
Qy 2761 AATTTTCAGCTTCAGACAGCAAGAGATGATTTTATGACGGGTAGCGACGCAATGCCATC 2820
Db 2761 AATTTTCAGCTTCAGACAGCAAGAGATGATTTTATGACGGGTAGCGACGCAATGCCATC 2820
Qy 2821 AATTCAGCTTCAGCAATATCCATTCCTGGCGTAAATGTCACCTTGGTGGAACCAATTC 2880
Db 2821 AATTCAGCTTCAGCAATATCCATTCCTGGCGTAAATGTCACCTTGGTGGAACCAATTC 2880
Qy 2881 AGCAGAGCAATACGGGGAATATTTACTATCGAGAGAGCAAAATGTTACGCTAGAAGCC 2940
Db 2881 AGCAGAGCAATACGGGGAATATTTACTATCGAGAGAGCAAAATGTTACGCTAGAAGCC 2940
Qy 2941 AATAAGCCGCCCTAATCAGCAAAACATAGAGGATAGAGTTATATAAATGTCGAGCTTGGC 3000
Db 2941 AATAAGCCGCCCTAATCAGCAAAACATAGAGGATAGAGTTATATAAATGTCGAGCTTGGC 3000
Qy 3001 GTTAATGGGAGTTTAAATTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTATT 3060
Db 3001 GTTAATGGGAGTTTAAATTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTATT 3060
Qy 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACACCGGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACACCGGCAATTTT 3120
Qy 3121 ACCAATAATGGCATCGCGAAATTAATATACACAGAGAGTGGTAAACCTTGGCAATGTT 3180
Db 3121 ACCAATAATGGCATCGCGAAATTAATATACACAGAGAGTGGTAAACCTTGGCAATGTT 3180
Qy 3181 ACCAATGATGGTATTAACATTTACCTACAGCTTAAGGCAACCAAGAGCAATCATC 3240
Db 3181 ACCAATGATGGTATTTAACAATTTACCTACAGCTTAAGGCAACCAAGAGCAATCATC 3240
Qy 3241 GGGGGAGATATATCAACAAAGAGAGCTTAAATTTACACAGAGTATATATGATGCT 3300
Db 3241 GGGGGAGATATATCAACAAAGAGAGCTTAAATTTACACAGAGTATATATGATGCT 3300
Qy 3301 GAAATCCAAATTTGGCGCAATATCTCCGCAAAAGAGGCAACCTCAGGATTTCTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGCAATATCTCCGCAAAAGAGGCAACCTCAGGATTTCTCCGAT 3360
Qy 3361 AAAATTAATATCAACCAACAGATTAACATCAAAAGGGTATTCGATGGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATATCAACCAACAGATTAACATCAAAAGGGTATTCGATGGAGAGGACTCTAGT 3420
Qy 3421 TCAGATGGCAAGTAATGCCAAGCTAATTAATTAACCAAGAAATTTGAAATGACAGAA 3480
Db 3421 TCAGATGGCAAGTAATGCCAAGCTAATTAATTAACCAAGAAATTTGAAATGACAGAA 3480
Qy 3481 GACCTTAAGTATTTAGGTTTCAATTAAGAGAGATTTACAGCAAGAGATGGTATGAGATTTA 3540
Db 3481 GACCTTAAGTATTTAGGTTTCAATTAAGAGAGATTTACAGCAAGAGATGGTATGAGATTTA 3540
Qy 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCGGAAGCAAAACAGTAACCTTTAAC 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCGGAAGCAAAACAGTAACCTTTAAC 3600
Qy 3601 AATGTTAAGATTTCAAAATCTCTGCTACGCTACATGTTGACACTAAATAGCAAGAGTG 3660
Db 3601 AATGTTAAGATTTCAAAATCTCTGCTACGCTACATGTTGACACTAAATAGCAAGAGTG 3660
Qy 3661 AAAACATCTAGCAGCAATGGGGAGCTGAAAGCAATAGCAGCAAGATACCGCTTAAT 3720
Db 3661 AAAACATCTAGCAGCAATGGGGAGCTGAAAGCAATAGCAGCAAGATACCGCTTAAT 3720
Qy 3721 ATTACTGCAAAAATGTAGAAGTAACAAAGATATTTCTCTCAAAACAGTAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAAGTAACAAAGATATTTCTCTCAAAACAGTAATATC 3780

Qy 3781 ACCCGTCGGAAAAGGTTACCACACAGCAGGCTGACCACTTAACGCAACAAATGCAAA 3840
Db 3781 ACCCGTCGGAAAAGGTTACCACACAGCAGGCTGACCACTTAACGCAACAAATGCAAA 3840
Qy 3841 GCAAGTATTACAACCAACAGAGTGATATCAGCGGTACAGTTTCGGGTACACGCTAAGT 3900
Db 3841 GCAAGTATTACAACCAACAGAGTGATATCAGCGGTACAGTTTCGGGTACACGCTAAGT 3900
Qy 3901 GTTAGCGCGACTGGTGATTTTAAACACTTAAATTCGGGCTCAAAATTTGAGCGAAATCGGT 3960
Db 3901 GTTAGCGCGACTGGTGATTTTAAACACTTAAATTCGGGCTCAAAATTTGAGCGAAATCGGT 3960
Qy 3961 GAGCTAATGTAAACAGTGCAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGCTAATGTAAACAGTGCAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATACGGTA 4020
Qy 4021 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATPGCACA 4080
Db 4021 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATPGCACA 4080
Qy 4081 GAAGGAGCTGCAACCTTTAAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTTAAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
Qy 4141 ATCAGCTTCAACTTAAGGTCAGGTAGACCTTTGGCTCAGAATGGTAGCATCGCAGGAAGC 4200
Db 4141 ATCAGCTTCAACTTAAGGTCAGGTAGACCTTTGGCTCAGAATGGTAGCATCGCAGGAAGC 4200
Qy 4201 ATTAATGCTGCTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Db 4201 ATTAATGCTGCTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Qy 4261 GATATTAAGCAACACAGCGGCACTTTGGTTATTACGCAAAAGATGCTAAGCTAATGCT 4320
Db 4261 GATATTAAGCAACACAGCGGCACTTTGGTTATTACGCAAAAGATGCTAAGCTAATGCT 4320
Qy 4321 GATGATCAGGTGATGATGACAGAGTGAATGAGCTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATGATCAGGTGATGATGACAGAGTGAATGAGCTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Qy 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCTGCGGATTTAAACACAGTAAATGGGTGA 4440
Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCTGCGGATTTAAACACAGTAAATGGGTGA 4440
Qy 4441 AATATCATTTTCAAAAGATGTTAGAAACACTGTCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500
Db 4441 AATATCATTTTCAAAAGATGTTAGAAACACTGTCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500
Qy 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAAAGATTAATTTGAAGCGAAACGCGCTT 4560
Db 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAAAGATTAATTTGAAGCGAAACGCGCTT 4560
Qy 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAAACTTTGGTGAAGT 4620
Db 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAAACTTTGGTGAAGT 4620
Qy 4621 GCTGTACGTTTCTTGGGCAAAATTAATCAATTAATCAATTAATCAATTAATTAATTAAT 4680
Db 4621 GCTGTACGTTTCTTGGGCAAAATTAATCAATTAATCAATTAATTAATTAATTAATTAAT 4680
Qy 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGGTCTTCTCAAGTGAAT 4740
Db 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGGTCTTCTCAAGTGAAT 4740
Qy 4741 GGGCAGCAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAAGTAATGACAAG 4800
Db 4741 GGGCAGCAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAAGTAATGACAAG 4800
Qy 4801 GTAGATTTCAATCTCTGCAATGAAGTCAATTTATTTCTGATTTATTTCTGTTGGGTAAA 4860
Db 4801 GTAGATTTCAATCTCTGCAATGAAGTCAATTTATTTCTGATTTATTTCTGTTGGGTAAA 4860

QY 4861 GTTCAGTACGGGCTTTACCCATCTGTAAAAAATACGAGAAATACAAATAAAGTATTTT 4920
 Db |||||||
 QY 4861 GTTCAGTACGGGCTTTACCCATCTGTAAAAAATACGAGAAATACAAATAAAGTATTTT 4920
 Db |||||||
 QY 4921 AACAGGTTATTATTATG 4937
 Db |||||||
 QY 4921 AACAGGTTATTATTATG 4937
 Db |||||||
 RESULT 3
 AAAS2197
 ID AAAS2197 standard; DNA; 4940 BP.
 AC AAAS2197;
 XX
 DT 11-SEP-2000 (first entry)
 DE
 XX
 DE Haemophilus influenzae strain 12 hmw2A gene, SEQ ID NO:70.
 XX
 KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis; ds.
 XX
 OS Haemophilus influenzae strain 12.
 XX
 FH Key Location/Qualifiers
 FT CDS 352..4785
 FT /*tag= a
 FT /product= "Haemophilus influenzae strain 12 hmw2A
 FT protein"
 XX
 PN WO200020609-A2.
 XX
 PD 13-APR-2000.
 XX
 XX 07-OCT-1999; 99WO-CA00938.
 PF
 XX 07-OCT-1998; 98US-0167568.
 PR
 XX 08-DEC-1998; 98US-0206942.
 PR
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 XX
 DR WPI; 2000-303789/26.
 DR P-PSDB; AAB01848.
 XX
 PT Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans.
 XX
 PS Example 16; Fig 29A-N; 307pp; English.
 XX
 CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMW. The invention also discloses hmwA genes (AAAS2175-A52198)
 CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
 CC strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in

CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an hmwA gene from a non-typeable strain of H.
 CC influenzae.
 XX

SQ Sequence 4940 BP; 1729 A; 948 C; 1011 G; 1252 T; 0 Other;

Query Match 99.7%; Score 4924; DB 21; Length 4940;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4937; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
 QY 1 TAAATATACAAGATAATAAATAAATCAAGATTTTCTGTGATGACAAACAATAATACAA 60
 Db |||||||
 QY 1 TAAATATACAAGATAATAAATAAATCAAGATTTTCTGTGATGACAAACAATAATACAA 60
 Db |||||||
 QY 61 CACCTTTTTCGAGTCTATATGCAAAATTTTAAAAAATAGTATAATCCGCATATAA 120
 Db |||||||
 QY 61 CACCTTTTTCGAGTCTATATGCAAAATTTTAAAAAATAGTATAATCCGCATATAA 120
 Db |||||||
 QY 121 AATGATATAATCTTTTCATCTTTTCATCTTTTAACTCTTCATCTTTTCATCTTTTCAT 180
 Db |||||||
 QY 121 AATGATATAATCTTTTCATCTTTTCATCTTTTAACTCTTTTCATCTTTTCATCTTTTCAT 180
 Db |||||||
 QY 181 GTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTT 240
 Db |||||||
 QY 181 GTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTT 240
 Db |||||||
 QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 Db |||||||
 QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 Db |||||||
 QY 301 GAAGCAAAATGATAAAGTAAATTTAATTTGTTCAACTTAGGAGAGAAATATGAACAAG 360
 Db |||||||
 QY 301 GAAGCAAAATGATAAAGTAAATTTAATTTGTTCAACTTAGGAGAGAAATATGAACAAG 360
 Db |||||||
 QY 361 ATATATCGTCTCAAAATTCAGCAAAACCGCTTCCCGTATGTTACTATCTTTAGGTGTAA 420
 Db |||||||
 QY 361 ATATATCGTCTCAAAATTCAGCAAAACCGCTTCCCGTATGTTACTATCTTTAGGTGTAA 420
 Db |||||||
 QY 421 CGGGTGTGACCATTCACAGAAAAAGGCTTCCCGTATGTTACTATCTTTAGGTGTAA 480
 Db |||||||
 QY 421 CGGGTGTGACCATTCACAGAAAAAGGCTTCCCGTATGTTACTATCTTTAGGTGTAA 480
 Db |||||||
 QY 481 CACTTAGCGTTAAAGCCACTTTCCCGTATGTTACTATCTTTAGGTGTAA 540
 Db |||||||
 QY 481 CACTTAGCGTTAAAGCCACTTTCCCGTATGTTACTATCTTTAGGTGTAA 540
 Db |||||||
 QY 541 CAATCTGTTTATAGCAAGCGCTTACAAGGAATGATGTAGTACAGGACACAGCCACTATG 600
 Db |||||||
 QY 541 CAATCTGTTTATAGCAAGCGCTTACAAGGAATGATGTAGTACAGGACACAGCCACTATG 600
 Db |||||||
 QY 601 CAAGTAGATGTTAAATAAACCAATTCGCGACAGCTGTTGAGCTATCATTAATTTGAAA 660
 Db |||||||
 QY 601 CAAGTAGATGTTAAATAAACCAATTCGCGACAGCTGTTGAGCTATCATTAATTTGAAA 660
 Db |||||||
 QY 661 CAATTTAAACATCGACCAAAATGAATGTGCGAGTTTTTACAAGAAAAACAACAACCTCGCC 720
 Db |||||||
 QY 661 CAATTTAAACATCGACCAAAATGAATGTGCGAGTTTTTACAAGAAAAACAACAACCTCGCC 720
 Db |||||||
 QY 721 GTATTCAACCGTGTGTACATCTTAACCAATCTCCCAATTTAAAGGATTTTAGATTCTAAC 780
 Db |||||||
 QY 721 GTATTCAACCGTGTGTACATCTTAACCAATCTCCCAATTTAAAGGATTTTAGATTCTAAC 780
 Db |||||||
 QY 781 GCACAAGTCTTTTAAATAACCAACCAATTTGATATCAATAGTAAGACGCAATTTATTAA 840
 Db |||||||
 QY 781 GCACAAGTCTTTTAAATAACCAACCAATTTGATATCAATAGTAAGACGCAATTTATTAA 840
 Db |||||||
 QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAAACAACAAGCGCGTAAT 900
 Db |||||||

Db 841 ACTAATGGCTTTACCGCTTCTACGCTAGACATTTCTAACCAAAACATCAAGCGCGTAAAT 900
QY 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACCGTTTAAAT 960
Db 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACCGTTTAAAT 960
QY 961 ACTGTCGGTAAAGACGCGAGCTTAAATCTTATTTGGTGGCAAGTGAAGAAACAGAGGTG 1020
Db 961 ACTGTCGGTAAAGACGCGAGCTTAAATCTTATTTGGTGGCAAGTGAAGAAACAGAGGTG 1020
QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTCGCAGGCGCAAAATACCAATCAGCGAT 1080
Db 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTCGCAGGCGCAAAATACCAATCAGCGAT 1080
QY 1081 ATATAAACCAACCACTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140
Db 1081 ATATAAACCAACCACTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140
QY 1141 GCGCATTTTTCGAAGCGGTACATTAATGTCGCTGCCACTATTCGAAACCAA 1200
Db 1141 GCGCATTTTTCGAAGCGGTACATTAATGTCGCTGCCACTATTCGAAACCAA 1200
QY 1201 GGTAACTTCTGCTGATTTCTGTAAGCAAAAGATAAAGCGGCAATTTGTTCTTCCGCC 1260
Db 1201 GGTAACTTCTGCTGATTTCTGTAAGCAAAAGATAAAGCGGCAATTTGTTCTTCCGCC 1260
QY 1261 AAAGAGGTGAAGCGGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
Db 1261 AAAGAGGTGAAGCGGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
QY 1321 GCGAGCTGATGATTACAGCGGATAAAGTACATTTAAAAACAGGTGCGATTTCGACCTT 1380
Db 1321 GCGAGCTGATGATTACAGCGGATAAAGTACATTTAAAAACAGGTGCGATTTCGACCTT 1380
QY 1381 TCAGTAAAGAAAGGGGAGAACTTACCTTGGCGGTGACAGCGCGCGAAGGTAAAGAAC 1440
Db 1381 TCAGTAAAGAAAGGGGAGAACTTACCTTGGCGGTGACAGCGCGCGAAGGTAAAGAAC 1440
QY 1441 GGCATTCATTTAGCAAAACCACTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1500
Db 1441 GGCATTCATTTAGCAAAACCACTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1500
QY 1501 AAAGAAAGCGGAGCGCTATTGTTGGGCGATATTGCGTTAATTTAGCGCAATATT 1560
Db 1501 AAAGAAAGCGGAGCGCTATTGTTGGGCGATATTGCGTTAATTTAGCGCAATATT 1560
QY 1561 AACGCTCAAGGTAGTGGTATATCTGTAANACC --- GGTGTTTGTGGAGACATCGGG 1617
Db 1561 AACGCTCAAGGTAGTGGTATATCTGTAANACC --- GGTGTTTGTGGAGACATCGGG 1620
QY 1618 CATTTATTCATTTGACAGCAATGCAATTTGTAACAAAGAGTGGTCTAGACCTT 1677
Db 1621 CATTTATTCATTTGACAGCAATGCAATTTGTAACAAAGAGTGGTCTAGACCTT 1680
QY 1678 GATGATGAACAAATTAAGCGGAGACCCCTTCGCAATATACCGGTATTAATGATGAA 1737
Db 1681 GATGATGAACAAATTAAGCGGAGACCCCTTCGCAATATACCGGTATTAATGATGAA 1740
QY 1738 TTCCCAACAGGACCGGTGAAGCAAGCGACCCCTTAAAAAAATAGCGAATCAAAACAG 1797
Db 1741 TTCCCAACAGGACCGGTGAAGCAAGCGACCCCTTAAAAAAATAGCGAATCAAAACAG 1800
QY 1798 CTAAACCAATCAACTATTTCAAATTTATCTGAAAAAGCGCTGGACAAATGAATTAACGGCA 1857
Db 1801 CTAAACCAATCAACTATTTCAAATTTATCTGAAAAAGCGCTGGACAAATGAATTAACGGCA 1860
QY 1858 TCAAGAAACTTACCGTAAATAGCTCAATCAATCGGAGCAACTCCACATTAATTC 1917
Db 1861 TCAAGAAACTTACCGTAAATAGCTCAATCAATCGGAGCAACTCCACATTAATTC 1920
QY 1918 CATAGTAAAGGTGAGCGTGGCGAGCGGTTCAGATTGAGATATTACTTCTAAGGC 1977
Db 1921 CATAGTAAAGGTGAGCGTGGCGAGCGGTTCAGATTGAGATATTACTTCTAAGGC 1980

QY 1978 GGAATTTAACCAATTTATCTGCGGATGGTTGATGTTTCATAAAAAATATTACGCTTGAT 2037
Db 1981 GGAATTTAACCAATTTATCTGCGGATGGTTGATGTTTCATAAAAAATATTACGCTTGAT 2040
QY 2038 CAGGCTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGTGGAAATTAACAAAGCA 2097
Db 2041 CAGGCTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGTGGAAATTAACAAAGCA 2100
QY 2098 CCGACGCGGCAATCTTAAATTTCTCCCGCAGGCACTGTAACCATTTACAGGAGAGGA 2157
Db 2101 CCGACGCGGCAATTTCTTAAATTTCTTAAACGGAAGGTTAAAGTCTGAATATCAT 2160
QY 2158 AAAGATTTTCAGGCTTAAACAGTATCTTAAACGGAAGGTTAAAGTCTGAATATCAT 2217
Db 2161 AAAGATTTTCAGGCTTAAACAGTATCTTAAACGGAAGGTTAAAGTCTGAATATCAT 2220
QY 2218 TCATCAGTGAATTAATTTAAACCCACAATCTTAGTGCAACATTTAACTATCTGGGAATATA 2277
Db 2221 TCATCAGTGAATTAATTTAAACCCACAATCTTAGTGCAACATTTAACTATCTGGGAATATA 2280
QY 2278 ACAATTTAACCAAACTACGAGAAAGAACCTCGTATTGGCAAAACGAGCATGATTCGCAC 2337
Db 2281 ACAATTTAACCAAACTACGAGAAAGAACCTCGTATTGGCAAAACGAGCATGATTCGCAC 2340
QY 2338 TGAAGCTCAGTGTCTTAAATCTTAGAGCAGGCGCAAAATTTTACCTTTTAAATATCAT 2397
Db 2341 TGAAGCTCAGTGTCTTAAATCTTAGAGCAGGCGCAAAATTTTACCTTTTAAATATCAT 2400
QY 2398 TCAAGCAATAGCAAAAGCTTAAACACACAGTATAGAAGCTCTGCGGGGTGAATTTTAAAC 2457
Db 2401 TCAAGCAATAGCAAAAGCTTAAACACACAGTATAGAAGCTCTGCGGGGTGAATTTTAAAC 2460
QY 2458 GCGTAAATGGCAACATGTCATTCATCTCAAAAGAGGAGCGAAAGTAAATTTCAAAATTA 2517
Db 2461 GCGTAAATGGCAACATGTCATTCATCTCAAAAGAGGAGCGAAAGTAAATTTCAAAATTA 2520
QY 2518 AAACCAACGAGAACATGACACAAACACCTTTACCAATTCGGTTTATAGCAATATC 2577
Db 2521 AAACCAACGAGAACATGACACAAACACCTTTACCAATTCGGTTTATAGCAATATC 2580
QY 2578 ACAGCCACTGGTGGGCTCTGTTTATGATATATATGCAACCATCTGCGCAGAGG 2637
Db 2581 ACAGCCACTGGTGGGCTCTGTTTATGATATATATGCAACCATCTGCGCAGAGG 2640
QY 2638 GCTGAGTTAAAAATGAGTGAATTAATCTCTAACGGCGCTAAATTTTACCTTAAATTC 2697
Db 2641 GCTGAGTTAAAAATGAGTGAATTAATCTCTAACGGCGCTAAATTTTACCTTAAATTC 2700
QY 2698 CATGTTCCGCGGATGACGCTTTTAAATCAACAAAGACTTTAACCATTAATGCAACCAAT 2757
Db 2701 CATGTTCCGCGGATGACGCTTTTAAATCAACAAAGACTTTAACCATTAATGCAACCAAT 2760
QY 2758 TCAATTTACGCTTCAGACAGACGAAAGATGATTTTATGACGGGTACGCAACCAATGCC 2817
Db 2761 TCAATTTACGCTTCAGACAGACGAAAGATGATTTTATGACGGGTACGCAACCAATGCC 2820
QY 2818 ATCAATTTCAACCTACACATATCCATCTGCGCGGTAAATGTCACCTTTGTTGGGACAAAC 2877
Db 2821 ATCAATTTCAACCTACACATATCCATCTGCGCGGTAAATGTCACCTTTGTTGGGACAAAC 2880
QY 2878 TCAAGCAGCAGATTTACGGGAATTTACTATCGAAGAGCAGCAAAATGTTTACGCTAGAA 2937
Db 2881 TCAAGCAGCAGATTTACGGGAATTTACTATCGAAGAGCAGCAAAATGTTTACGCTAGAA 2940
QY 2938 GCCAATAACGCCCTTAATCAGCAAAACATAAGGATAGAGTTTAAACTTGGCAGCTG 2997
Db 2941 GCCAATAACGCCCTTAATCAGCAAAACATAAGGATAGAGTTTAAACTTGGCAGCTG 3000
QY 2998 CTCGTTAATGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTAAAGGCAATCTCACT 3057
Db 3001 CTCGTTAATGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTAAAGGCAATCTCACT 3060

|||||
Db 1681 GATGTAACAAATTTGAGCCGGAAGACCCCTTCCGAATAATACCGGTATAAATGATGAATTC 1740
QY 1741 CCAACAGGACCGGTGAAGCAAGCGACCCCTAAAAAAATAGCGAACTCAAAAACACGCTA 1800
Db 1741 CCAACAGGACCGGTGAAGCAAGCGACCCCTAAAAAAATAGCGAACTCAAAAACACGCTA 1800
QY 1801 ACCAATACAACTATTTCAAATATTCTGAANAAGCGCTGGACAAATGAATATAACGGCATCA 1860
Db 1801 ACCAATACAACTATTTCAAATATTCTGAANAAGCGCTGGACAAATGAATATAACGGCATCA 1860
QY 1861 AGAAAACCTTACCCTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920
Db 1861 AGAAAACCTTACCCTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920
QY 1921 AGTAAAGTACGCGTGGCGGAGCGCTTCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980
Db 1921 AGTAAAGTACGCGTGGCGGAGCGCTTCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980
QY 1981 AATTTAAACCATTTATCTGGCGGATGGCTTGATGTTTCATAAAATATTACGGTTGATCAG 2040
Db 1981 AATTTAAACCATTTATCTGGCGGATGGCTTGATGTTTCATAAAATATTACGGTTGATCAG 2040
QY 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATAACAAAGCACGC 2100
Db 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATAACAAAGCACGC 2100
QY 2101 GACGCGCAATGCTAAATTTGTGCCCGACGCTGTAAACATTTACAGGAGAGGGA 2160
Db 2101 GACGCGCAATGCTAAATTTGTGCCCGACGCTGTAAACATTTACAGGAGAGGGA 2160
QY 2161 GATTTAGGGCTAACAAAGCTATCTTTAAACGGAAGCGGTAAAGGTCTCAATATCATTTCA 2220
Db 2161 GATTTAGGGCTAACAAAGCTATCTTTAAACGGAAGCGGTAAAGGTCTCAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGGCAATTAACATATCTGGAATATAACA 2280
Db 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGGCAATTAACATATCTGGAATATAACA 2280
QY 2281 ATTAACCAACTACGAGAAACACCTCTGTTTGGCAACAGCGCATGATTCGCACCTGG 2340
Db 2281 ATTAACCAACTACGAGAAACACCTCTGTTTGGCAACAGCGCATGATTCGCACCTGG 2340
QY 2341 AACCTAGTCTTAACTTAGACACAGCGCAATTTTACCTTTTAAATACATTTCA 2400
Db 2341 AACCTAGTCTTAACTTAGACACAGCGCAATTTTACCTTTTAAATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACACAGATAGAAAGCTCTGAGGGGTGAATTTTAAACGGC 2460
Db 2401 AGCAATAGCAAGGCTTAAACACACAGATAGAAAGCTCTGAGGGGTGAATTTTAAACGGC 2460
QY 2461 GTAATGGCAACATGTTCATTTCAATCTCAAAGAGGAGCGAAAGTTAATTTCAATTA 2520
Db 2461 GTAATGGCAACATGTTCATTTCAATCTCAAAGAGGAGCGAAAGTTAATTTCAATTA 2520
QY 2521 CCAACGAGAACATGAACACAGCAACCTTTTACCAATTCGGTTTTTAGCCAATATCA 2580
Db 2521 CCAACGAGAACATGAACACAGCAACCTTTTACCAATTCGGTTTTTAGCCAATATCA 2580
QY 2581 GCCACTGGTGGGCTCTGTTTTTTTGTATATATGCGCAACCATCTGCGAGAGGGCT 2640
Db 2581 GCCACTGGTGGGCTCTGTTTTTTTGTATATATGCGCAACCATCTGCGAGAGGGCT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGGCTCAATTTTACCTTAAATTTCCCAT 2700
Db 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGGCTCAATTTTACCTTAAATTTCCCAT 2700
QY 2701 GTTCGGGCGATGACGCTTTTAAAAATCAACAAAGCTTAACCAATAATGCAACCAATTTCA 2760
Db 2701 GTTCGGGCGATGACGCTTTTAAAAATCAACAAAGCTTAACCAATAATGCAACCAATTTCA 2760
QY 2761 AATTTACGCTCAGACAGCAAGATGATTTTATGAGGGGTACGAGCAATGCCATC 2820
|||||

Db 2761 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGTACGACCAATGCCATC 2820
QY 2821 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGTACGACCAATGCCATC 2880
Db 2821 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGTACGACCAATGCCATC 2880
QY 2881 AGCAGCAGCATTTACGGGGAATATTACTATCTGAGAAAGCAGCAATCTTACGCTAGAGCC 2940
Db 2881 AGCAGCAGCATTTACGGGGAATATTACTATCTGAGAAAGCAGCAATCTTACGCTAGAGCC 2940
QY 2941 AATAAGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGGCTC 3000
Db 2941 AATAAGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGGCTC 3000
QY 3001 GTTAATGGGAGTTTAACTTAACTGGCGAAATGCGAGATTTAAAGGCAATCTCACTATT 3060
Db 3001 GTTAATGGGAGTTTAACTTAACTGGCGAAATGCGAGATTTAAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAAATATCACCGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAAATATCACCGCAATTTT 3120
QY 3121 ACCAATATGGCAGTGGCGAAATTAATATACACAGGAGTGGTAAACTTGGCAATGTT 3180
Db 3121 ACCAATATGGCAGTGGCGAAATTAATATACACAGGAGTGGTAAACTTGGCAATGTT 3180
QY 3181 ACCAATGATGTTGATTTAAACATTTACCCTACGCTTAACGCAACCAAGAGCATCTC 3240
Db 3181 ACCAATGATGTTGATTTAAACATTTACCCTACGCTTAACGCAACCAAGAGCATCTC 3240
QY 3241 GCGGAGATTAATCAACAAAGAGTAAATATTTACAGACAGTAATTAATGATGCT 3300
Db 3241 GCGGAGATTAATCAACAAAGAGTAAATATTTACAGACAGTAATTAATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGAGGCAACCTTCTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGAGGCAACCTTCTCCGAT 3360
QY 3361 AAAATTAATATCACCACAGATTAACATCAAAAGAGTAAATTTACAGACAGTAATTAATGATGCT 3420
Db 3361 AAAATTAATATCACCACAGATTAACATCAAAAGAGTAAATTTACAGACAGTAATTAATGATGCT 3420
QY 3421 TCAGATCGCAAGTAAATGCGCAACCTTAATTTAAACCAAGAAATTTGAGAGAA 3480
Db 3421 TCAGATCGCAAGTAAATGCGCAACCTTAATTTAAACCAAGAAATTTGAGAGAA 3480
QY 3481 GACCTAAGTATTTTACAGGTTTCAATTAAGCAGAGATTTACAGCAAGATTTAGAGATTTA 3540
Db 3481 GACCTAAGTATTTTACAGGTTTCAATTAAGCAGAGATTTACAGCAAGATTTAGAGATTTA 3540
QY 3541 ACTATTCGCAACATTAATGACGGTAAACAGCGGTGCGGAAGCAAAACAGTAACTTTTAA 3600
Db 3541 ACTATTCGCAACATTAATGACGGTAAACAGCGGTGCGGAAGCAAAACAGTAACTTTTAA 3600
QY 3601 AATGTTAAAGATTTCAAAATCTCTGCTGACGCTCAGATTTGACACTTAATAGCAAGTG 3660
Db 3601 AATGTTAAAGATTTCAAAATCTCTGCTGACGCTCAGATTTGACACTTAATAGCAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGCGGAGCTGAAAGCAATAGCGCAACGATACCGGCTTAACT 3720
Db 3661 AAAACATCTAGCAGCAATGCGGAGCTGAAAGCAATAGCGCAACGATACCGGCTTAACT 3720
QY 3721 ATTACTGCAAAATGATAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780
Db 3721 ATTACTGCAAAATGATAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780
QY 3781 ACCGCTCGGAAAGGTTTACCACAGAGGCTGCGACCATTAACGCAACAAATGGCAAA 3840
Db 3781 ACCGCTCGGAAAGGTTTACCACAGAGGCTGCGACCATTAACGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACAAACCAACAGGTTGATATCAGCGGTACGATTTCCCGTTAACCGGTAAGT 3900
Db 3841 GCAAGTATTACAAACCAACAGGTTGATATCAGCGGTACGATTTCCCGTTAACCGGTAAGT 3900

Qy	3901	GTTAGCGGACTGGTGATTTAAACACTAAATCCGGCTCAAAATTTGAAGCGAAATCCGGT	3960
Db	3901	GTTAGCGGACTGGTGATTTAAACACTAAATCCGGCTCAAAATTTGAAGCGAAATCCGGT	3960
Qy	3961	GAGGCTAATGTAACAAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAACGGTA	4020
Db	3961	GAGGCTAATGTAACAAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAACGGTA	4020
Qy	4021	ATGTGTACGGCAACGCTGGCGATTTTAAACAGTTTGGGAATGGCGCAGAAATTAATCGGACA	4080
Db	4021	ATGTGTACGGCAACGCTGGCGATTTTAAACAGTTTGGGAATGGCGCAGAAATTAATCGGACA	4080
Qy	4081	GAAGGAGCTGCAACCTTAAACCGCAACAGGGAATACCTTGACTACTGAACGCGTTCCTAGC	4140
Db	4081	GAAGGAGCTGCAACCTTAAACCGCAACAGGGAATACCTTGACTACTGAACGCGTTCCTAGC	4140
Qy	4141	ATCACCTTCAACTAAGGGTCAGGTAGACCTCTTTGGCTCTCAGAATGGTGTAGTATCCAGGAAGC	4200
Db	4141	ATCACCTTCAACTAAGGGTCAGGTAGACCTCTTTGGCTCTCAGAATGGTGTAGTATCCAGGAAGC	4200
Qy	4201	ATTTAATGCTGCTAAATGTGACATTTAAATACTACAGGCACCTTAACCCACCGTGGCAGGCTCG	4260
Db	4201	ATTTAATGCTGCTAAATGTGACATTTAAATACTACAGGCACCTTAACCCACCGTGGCAGGCTCG	4260
Qy	4261	GATATTTAAAGCAACACAGCGGCACCTTGGTGTATTAAACGCAAAAGATGCTAAGCTAAATGCT	4320
Db	4261	GATATTTAAAGCAACACAGCGGCACCTTGGTGTATTAAACGCAAAAGATGCTAAGCTAAATGCT	4320
Qy	4321	GATGCATCAGGTGATAGTACAGAACTGAATCAGCTCAACCGCAACGCGCTCTGGTAGTGTCG	4380
Db	4321	GATGCATCAGGTGATAGTACAGAACTGAATCAGCTCAACCGCAACGCGCTCTGGTAGTGTCG	4380
Qy	4381	ACTGCGGCAACCTCAAGCAGTGTGAATATCACTCTGGGATTTAAACACAGTAATGGGTTA	4440
Db	4381	ACTGCGGCAACCTCAAGCAGTGTGAATATCACTCTGGGATTTAAACACAGTAATGGGTTA	4440
Qy	4441	AAATATCATTTTCTGAAAGATGGTGAAGAACACTGTGCGCTTAAGAGCAAGGAAATTGAGTGT	4500
Db	4441	AAATATCATTTTCTGAAAGATGGTGAAGAACACTGTGCGCTTAAGAGCAAGGAAATTGAGTGT	4500
Qy	4501	AAATATATCCAGCGCAGGTGTAGCAAGTGTAGAAAGCTTAATTGAAGCGAAACCGCTCTT	4560
Db	4501	AAATATATCCAGCGCAGGTGTAGCAAGTGTAGAAAGCTTAATTGAAGCGAAACCGCTCTT	4560
Qy	4561	GAAAGCTTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACCTTGGTGTAAAGT	4620
Db	4561	GAAAGCTTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACCTTGGTGTAAAGT	4620
Qy	4621	GCTGTAGCTTTTGTGAGCCAAATATACAAATACAGTCAATACACAAAATGAATTTACA	4680
Db	4621	GCTGTAGCTTTTGTGAGCCAAATATACAAATACAGTCAATACACAAAATGAATTTACA	4680
Qy	4681	ACCAGACCGTCAAGTCAAGTGATAATTTCTGAAGTGAAGCGGTGTTTCTCAAGTGGTAAT	4740
Db	4681	ACCAGACCGTCAAGTCAAGTGATAATTTCTGAAGTGAAGCGGTGTTTCTCAAGTGGTAAT	4740
Qy	4741	GGCGCACGAGTATGTACCAATGTGCTGACGATGGACAGCGCTAGTCAGTAATTTGACAAG	4800
Db	4741	GGCGCACGAGTATGTACCAATGTGCTGACGATGGACAGCGCTAGTCAGTAATTTGACAAG	4800
Qy	4801	GTGATATTTCACTCTGCAATGAAGTCAATTTATTTTTCGTTATTTTACTGTCTGGGTTAAA	4860
Db	4801	GTGATATTTCACTCTGCAATGAAGTCAATTTATTTTTCGTTATTTTACTGTCTGGGTTAAA	4860
Qy	4861	GTTTCAGTACGGGCTTTTACCCATCTCTGTAATAAATTACCGAGAAATCAATAAGTATTTTT	4920
Db	4861	GTTTCAGTACGGGCTTTTACCCATCTCTGTAATAAATTACCGAGAAATCAATAAGTATTTTT	4920
Qy	4921	AAACAGGTTATTATTATG 4937	
Db	4921	AAACAGGTTATTATTATG 4937	

RESULT 5
AA049509 standard; DNA; 9323 BP.
ID AAQ49509;
XX AC AAQ49509;
XX DT 26-APR-1994 (first entry)
XX DE Gene cluster for high molecular weight protein 2 (HMW2).
XX HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae; gene cluster; ss.
XX OS Haemophilus influenzae.
XX XX
XX Key Location/Qualifiers
PH 792..5222
FT CDS /*tag= a
FT FT /product= "High molecular weight protein 2"
FT 5375..7009
CDS /*tag= b
FT FT /note= "One of a gene cluster for High molecular"
FT weight protein 2"
FT 7149..9098
CDS /tag= c
FT FT /note= "One of a gene cluster for High molecular"
FT weight protein 2"
FT FT
XX W09319090-A.
PV
XX 30-SEP-1993.
PD
XX 16-MAR-1993; 93WO-US02166.
PF
XX 16-MAR-1992; 92GB-0005704.
PR
XX (BARE/) BARENKAMP S J.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Barenkamp SJ;
XX WPT: 1993-320683/40.
DR P-PSDB; AAR41728, AAR41729, AAR41730.
XX High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
PS Claim 10; Figure 7; 100pp; English.
XX The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX Sequence 9323 BP; 3133 A; 1829 C; 1833 G; 2528 T; 0 other;
SQ
Query Match 97.8%; Score 4828.2; DB 14; Length 9323;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 4869; Conservative 0; Mismatches 68; Indels 0; Gaps
OY 1 TAAATACACAAGTAATAAATAAATAAATCAAGTTTTTCTGATGACAAACAACAAATTACAA 60
Db 441 TAAATACACAAGTAATAAATAAATAAATCAAGTTTTTCTGATGACAAACAACAAATTACAA 500
OY 61 CACCTTTTTCGAGTCATATATGCCAAATATTTTAAAAAATAGTATATATCGGCATATAA 120
Db 501 CACCTTTTTCGAGTCATATATGCCAAATATTTTAAAAAATAGTATATATCGGCATATAA 560
OY 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTAAFCFTTTCATCTTTTCATCTTTTCAT 180
Db 561 AATGGTATAATCTTTTCATCTTTTCATCTTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620

[illegible]

Qy 781 GGACAGCTCTTTTAAATCAACCCAAAGGGHATCACAAATAGGTAAAGACGCAATTTAAAC 840
Db 1221 GGACAGCTCTTTTAAATCAACCCAAATGGTATCACAAATAGGTAAAGACGCAATTTAAAC 1280
Qy 841 ACTAATGGCTTACGGCTTCTAGCTAGACATTTCTAAGCAAAACATCAAGGCGCGTAAT 900
Db 1281 ACTAATGGCTTACGGCTTCTAGCTAGACATTTCTAAGCAAAACATCAAGGCGCGTAAT 1340
Qy 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTCGTGAATCACGGTCTTAAAT 960
Db 1341 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTCGTGAATCACGGTCTTAAAT 1400
Qy 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGGTGGCAAGGTGAAACAGAGGCTGTG 1020
Db 1401 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGGTGGCAAGGTGAAACAGAGGCTGTG 1460
Qy 1021 ATTAGCCTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATTCACCATCAGCGAT 1080
Db 1461 ATTAGCCTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATTCACCATCAGCGAT 1520
Qy 1081 ATAAATAACCCAAACATTTACTTACAGCATTTCCCGCGCTGAAATGAGCGGTCAATCTG 1140
Db 1521 ATAAATAACCCAAACATTTACTTACAGCATTTCCCGCGCTGAAATGAGCGGTCAATCTG 1580
Qy 1141 GGCATATTTTGGCCAAAGGGGTAACATTAATGTCGCTGCTGCGCAGTATTCGAACCAA 1200
Db 1581 GGCATATTTTGGCCAAAGGGGTAACATTAATGTCGCTGCTGCGCAGTATTCGAACCAA 1640
Qy 1201 GGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAGCGCGCAATATTGTTCTTCCGCC 1260
Db 1641 GGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAGCGCGCAATATTGTTCTTCCGCC 1700
Qy 1261 AAAGAGGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1320
Db 1701 AAAGAGGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1760
Qy 1321 GGCAGCTGATGATTACAGGCGGATTAAGTGCATTTAAACAGAGTGCAGTTATCGACCTT 1380
Db 1761 GGCAGCTGATGATTAAGTGCATTTAAAGTGCATTTAAACAGAGTGCAGTTATCGACCTT 1820
Qy 1381 TCAGTTAAAGAGGGGGAACCTTACCTTCGCGGTGACAGCGCGCGGAGGTAAACAA 1440
Db 1821 TCAGTTAAAGAGGGGGAACCTTACCTTCGCGGTGACAGCGCGCGGAGGTAAACAA 1880
Qy 1441 GGCATTCATTTAGCAAAAGAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500
Db 1881 GGCATTCATTTAGCAAAAGAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1940
Qy 1501 AAAGAAAAGGGGACGCGCTATTGTTGGGCGGATATTGGGTTAAATGACGGCAATTT 1560
Db 1941 AAAGAAAAGGGGACGCGCTATTGTTGGGCGGATATTGGGTTAAATGACGGCAATTT 2000
Qy 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAAACCGGTGTTTGTGGAGACATCGGGGCAT 1620
Db 2001 AACGCTCAAGGTAGTGGTGATATCGCTAAAACCGGTGTTTGTGGAGACATCGGGGCAT 2060
Qy 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTCGTAGACCCCTGAT 1680
Db 2061 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTCGTAGACCCCTGAT 2120
Qy 1681 GATGTAACANTTGAAGCGGAGAGCCCTTCGCAATATACCGGTATTAATGATCAATTC 1740
Db 2121 GATGTAACANTTGAAGCGGAGAGCCCTTCGCAATATACCGGTATTAATGATCAATTC 2180
Qy 1741 CCAACAGGACCGGTGAAGCAAGGACCCCTTAAAAAATAGGGAATCAAAACACGCTA 1800
Db 2181 CCAACAGGACCGGTGAAGCAAGGACCCCTTAAAAAATAGGGAATCAAAACACGCTA 2240
Qy 1801 ACCAATACAACTATTTTCAAAATTTCTGAAAAACGCGCTGGACAAATGAATATTAACGGCATCA 1860
Db 2241 ACCAATACAACTATTTTCAAAATTTCTGAAAAACGCGCTGGACAAATGAATATTAACGGCATCA 2300

Qy 1861 AGAAACCTTACCGTTAATAGCTCAATCAATCGGAACCACTCCCATCTTAAATCTCCAT 1920
Db 2301 AGAAACCTTACCGTTAATAGCTCAATCAATCGGAACCACTCCCATCTTAAATCTCCAT 2360
Qy 1921 AGTAAAGTTCAGCTGGGAGGCGCTTCAGATTGATGAGATATTACTTCTTAAAGGCGGA 1980
Db 2361 AGTAAAGTTCAGCTGGGAGGCGCTTCAGATTGATGAGATATTACTTCTTAAAGGCGGA 2420
Qy 1981 AATTTAACCATTTATTCCTGGCGGATGGGTGATGTTTCATFAAAATATTACGCTTGATCAG 2040
Db 2421 AATTTAACCATTTATTCCTGGCGGATGGGTGATGTTTCATFAAAATATTACGCTTGATCAG 2480
Qy 2041 GGTCTTTTAAATATTACCGCGCTTCCGCTAGCTTTTGAAGGTGGAATAAACAAGACGCG 2100
Db 2481 GGTCTTTTAAATATTACCGCGCTTCCGCTAGCTTTTGAAGGTGGAATAAACAAGACGCG 2540
Qy 2101 GACGGGGCAAAATGCTAAAAATTCGCGCCAGGGCACTGTAAACCATTTACAGGAGGGA 2160
Db 2541 GACGGGGCAAAATGCTAAAAATTCGCGCCAGGGCACTGTAAACCATTTACAGGAGGGA 2600
Qy 2161 GATTTTCAGGCTTAAACAACGCTATCTTTAAACGGAACGGGTAAAAGGCTGTAATATCTTCA 2220
Db 2601 GATTTTCAGGCTTAAACAACGCTATCTTTAAACGGAACGGGTAAAAGGCTGTAATATCTTCA 2660
Qy 2221 TCAGTGAATTAATTTAACCCACATCTTACTGTCACAAATTAACATATCTCGGAATAAACA 2280
Db 2661 TCAGTGAATTAATTTAACCCACATCTTACTGTCACAAATTAACATATCTCGGAATAAACA 2720
Qy 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGGCATGTTGCGACTGG 2340
Db 2721 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGGCATGTTGCGACTGG 2780
Qy 2341 AACGTCACTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTATTAATACATTTCA 2400
Db 2781 AACGTCACTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTATTAATACATTTCA 2840
Qy 2401 AGCAATAGCAAAAGGCTTAAACAACAGATATAGAGCTCTGAGGGGTGAATTTTAAACGCG 2460
Db 2841 AGCAATAGCAAAAGGCTTAAACAACAGATATAGAGCTCTGAGGGGTGAATTTTAAACGCG 2900
Qy 2461 GTAAATGGCAACATGTCTTCAATCTCAAGAAAGAGGCGAAAGTTTAAATTTCAATTTAAA 2520
Db 2901 GTAAATGGCAACATGTCTTCAATCTCAAGAAAGAGGCGAAAGTTTAAATTTCAATTTAAA 2960
Qy 2521 CCAACGAGAACATGAACACAAACACCTTTTACCAATTCGCTTTTATAGCAATATCACA 2580
Db 2961 CCAACGAGAACATGAACACAAACACCTTTTACCAATTCGCTTTTATAGCAATATCACA 3020
Qy 2581 GCCACTGGTGGGCTCTGTTTGTGATATATATGCCAACCATTTCTGGCAGAGGGCT 2640
Db 3021 GCCACTGGTGGGCTCTGTTTGTGATATATATGCCAACCATTTCTGGCAGAGGGCT 3080
Qy 2641 GAGTTAAATGAGTGAATTAATATCTCTAACGGGCTTAATTTTACCTTTAAATTTCCCAT 2700
Db 3081 GAGTTAAATGAGTGAATTAATATCTCTAACGGGCTTAATTTTACCTTTAAATTTCCCAT 3140
Qy 2701 GTTCGGCGGATGACGCTTTTAAATAACAAGACTTAAACCAATAATGCAACCAATTTCA 2760
Db 3141 GTTCGGCGGATGACGCTTTTAAATAACAAGACTTAAACCAATAATGCAACCAATTTCA 3200
Qy 2761 AATTTTCAGCTTCAGACAGAGAAAGATGATTTTATGACGGGTACGCAACCAATGCGCATC 2820
Db 3201 AATTTTCAGCTTCAGACAGAGAAAGATGATTTTATGACGGGTACGCAACCAATGCGCATC 3260
Qy 2821 AATTTTCAGCTTCAGACATATCTCTGGGCGGTAAATGTCACCTTGTGGCAAAACCTCA 2880
Db 3261 AATTTTCAGCTTCAGACATATCTCTGGGCGGTAAATGTCACCTTGTGGCAAAACCTCA 3320
Qy 2881 AGCAGCAGCATTCGCGGATATTTACTATCGAGAAACAGCAAAATGTTAGCTTGAAGGC 2940
Db 3321 AGCAGCAGCATTCGCGGATATTTACTATCGAGAAACAGCAAAATGTTAGCTTGAAGGC 3380
Qy 2941 AATAACGCCCTTAATCAGCAAAACATTAAGGATAGAGTTTATAAACTTTGCGAGCTTGCTC 3000

```
|||||
Db 3381 AATAACGCCCTAATCAGCAAAACATAAGGATAGAGTTTAAACCTTGGCAGCTTGCCTC 3440
QY 3001 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCGAGATATTAAAGCAATCTCACTATT 3060
Db 3441 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCGAGATATTAAAGCAATCTCACTATT 3500
QY 3061 TCAGAAAGGCCACCTTTTAAAGGAAGAGACTAGAGATACCCCTAAATATCAACCGCAATTTT 3120
Db 3501 TCAGAAAGGCCACCTTTTAAAGGAAGAGACTAGAGATACCCCTAAATATCAACCGCAATTTT 3560
QY 3121 ACCAATTAATGGCACTGCCGAAATTAATATAACACAAAGGAGTGTAAACCTTGGCAATGTT 3180
Db 3561 ACCAATTAATGGCACTGCCGAAATTAATATAACACAAAGGAGTGTAAACCTTGGCAATGTT 3620
QY 3181 ACCAATGATGGTGATTTAAACATTTACCCTACCGCTTAACAGCAACCAAGACATCATC 3240
Db 3621 ACCAATGATGGTGATTTAAACATTTACCCTACCGCTTAACAGCAACCAAGACATCATC 3680
QY 3241 GCGGAGATATATCAACAAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300
Db 3681 GCGGAGATATATCAACAAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCT 3740
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCAAGATTTCTCCGAT 3360
Db 3741 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCAAGATTTCTCCGAT 3800
QY 3361 AAAATTAATATCAACAAACAGATACAAATCAAAAAGGATTTGATGGAGAGGACTCTAGT 3420
Db 3801 AAAATTAATATCAACAAACAGATACAAATCAAAAAGGATTTGATGGAGAGGACTCTAGT 3860
QY 3421 TCAGATGCCACAAGTAATGCCAACCTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480
Db 3861 TCAGATGCCACAAGTAATGCCAACCTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3920
QY 3481 GACCTAAAGTATTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATTTGAAATTTGACAGAA 3540
Db 3921 GACCTAAAGTATTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATTTGAAATTTGACAGAA 3980
QY 3541 ACTATTTGCAACAGTAATGACGGTTACAGCGGTGCGAAGCCAAACAGTAATCTTTTAAAC 3600
Db 3981 ACTATTTGCAACAGTAATGACGGTTACAGCGGTGCGAAGCCAAACAGTAATCTTTTAAAC 3660
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAACATGTCACACTTAATAGCAAGTG 3660
Db 4041 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAACATGTCACACTTAATAGCAAGTG 4100
QY 3661 AAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCAGCAACGATACCGGCTTAAT 3720
Db 4101 AAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCAGCAACGATACCGGCTTAAT 4160
QY 3721 ATTACTGCAAAAATGTAGAAGTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780
Db 4161 ATTACTGCAAAAATGTAGAAGTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 4220
QY 3781 ACCCGCTCGGAAAAGTTTACCACCACAGCGCTCGACCATTTAAGCAACAAATGGCAAA 3840
Db 4221 ACCCGCTCGGAAAAGTTTACCACCACAGCGCTCGACCATTTAAGCAACAAATGGCAAA 4280
QY 3841 GCAAGTATTACACCAACAAAGGTGATATACGGGTACGATTTCCGGTTAACACCGTAAGT 3900
Db 4281 GCAAGTATTACACCAACAAAGGTGATATACGGGTACGATTTCCGGTTAACACCGTAAGT 4340
QY 3901 GTTAGCGGACTGGTGTATTAAACCACTAAATCCGGCTCAAAAATTTAGAGCAAAATCGGGT 3960
Db 4341 GTTAGCGGACTGGTGTATTAAACCACTAAATCCGGCTCAAAAATTTAGAGCAAAATCGGGT 4400
QY 3961 GAGGCTAATGTAAACAGTGCACAGGTACAAATTTGCGGTACAAATTTCCGGTAAATACGGTA 4020
Db 4401 GAGGCTAATGTAAACAGTGCACAGGTACAAATTTGCGGTACAAATTTCCGGTAAATACGGTA 4460
QY 4021 AATGTTAGGCAACGCTGGGATTTACAGTTGGGATTTGGGATTTGGCGAGAAATTAATGCGACA 4080
|||||
```

```
Db 4461 AATGTTACGGCAAAACGCTGGCGATTAAACAGTTGGGAATGGCGAGAAATTAATGCGACA 4520
QY 4081 GAAGGAGCTGCAACCTTTAAACCGCAACAGGAATACCTTTGACTACTGAAGCCGTTCTAGC 4140
Db 4521 GAAGGAGCTGCAACCTTTAAACCGCAACAGGAATACCTTTGACTACTGAAGCCGTTCTAGC 4580
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGAGCTCTTGGCTCAGAAATGCTAGCATCGCAGGAAGC 4200
Db 4581 ATCACTTCAACTAAGGCTCAGGTAGAGCTCTTGGCTCAGAAATGCTAGCATCGCAGGAAGC 4640
QY 4201 ATTAATGCTCTTAATGTGACATTAATACCTACAGGCACCTTTAACCCCGTGGCAGGCTCG 4260
Db 4641 ATTAATGCTCTTAATGTGACATTAATACCTACAGGCACCTTTAACCCCGTGGCAGGCTCG 4700
QY 4261 GATATTAAAGCAACACCGCGCACCTTTGGTTATTAAACCAAAAGATGCTAAAGTAAATGGT 4320
Db 4701 GATATTAAAGCAACACCGCGCACCTTTGGTTATTAAACCAAAAGATGCTAAAGTAAATGGT 4760
QY 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACCGAAGCGGCTCTGTAGTGTG 4380
Db 4761 GATGCATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACCGAAGCGGCTCTGTAGTGTG 4820
QY 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGGTTA 4440
Db 4821 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGGTTA 4880
QY 4441 AATATCATTTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAAGGCAAGGAAATTTAGGTTG 4500
Db 4881 AATATCATTTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAAGGCAAGGAAATTTAGGTTG 4940
QY 4501 AAATATATCAGCGCACGCTGTAGCAAGTGTAGAAAGAGTAAATTCGAAGCAACCGCTCCTT 4560
Db 4941 AAATATATCAGCGCACGCTGTAGCAAGTGTAGAAAGAGTAAATTCGAAGCAACCGCTCCTT 5000
QY 4561 GAAAGAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACCTTGGTGAAGT 4620
Db 5001 GAAAGAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACCTTGGTGAAGT 5060
QY 4621 GCTGTACGTTTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAATGAAATTTACA 4680
Db 5061 GCTGTACGTTTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAATGAAATTTACA 5120
QY 4681 ACCAGACCTCAAGTCAAGTGAATAATTTCTGAAGSTAAAGCGCTGTTTCTCAAGTGGTAAT 4740
Db 5121 ACCAGACCTCAAGTCAAGTGAATAATTTCTGAAGSTAAAGCGCTGTTTCTCAAGTGGTAAT 5180
QY 4741 GCGCACGAGTATGTACCAATCTTCTGACGATGGACACCGCTAGTCAGTAAATGACAAG 4800
Db 5181 GCGCACGAGTATGTACCAATCTTCTGACGATGGACACCGCTAGTCAGTAAATGACAAG 5240
QY 4801 GTAGATTTTCATCCTGCAATGAAGTCAATTTATTTTCTGATTTTACTCTGTGGGTAAA 4860
Db 5241 GTAGATTTTCATCCTGCAATGAAGTCAATTTATTTTCTGATTTTACTCTGTGGGTAAA 5300
QY 4861 GTTCAAGTACGGCTTTACCCATCTTGTAAAAAATTCAGGAGAAATACAAATTAAGTATTTT 4920
Db 5301 GTTCAAGTACGGCTTTACCCATCTTGTAAAAAATTCAGGAGAAATACAAATTAAGTATTTT 5360
QY 4921 AACAGGTTATTATTATG 4937
Db 5361 AACAGGTTATTATTATG 5377
```

RESULT 7

AAA52198

ID AAA52198 standard; DNA; 3263 BP.

XX AAA52198;

XX AC

XX 11-SEP-2000 (first entry)

XX DE

XX DNA encoding Haemophilus influenzae strain 12 mature hmw2A, SEQ ID NO:72.


```

ID AAQ49506 standard; DNA; 5116 BP.
XX
AC AAQ49506;
XX
XX 26-APR-1994 (first entry)
XX
XX Sequence encoding high molecular weight protein 1 (HWM1).
XX
XX HWM; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae; ss.
XX
XX Haemophilus influenzae.
XX
XX Key Location/Qualifiers
XX CDS 351..4961
XX FT /*tag= a
XX FT /product= High molecular weight protein 1
XX FT misc_difference 360..362
XX FT /*tag= b
XX FT /transl_except= CTA encodes Isoleucine.
XX FT misc_difference 642..644
XX FT /*tag= c
XX FT /transl_except= GAU encodes Alanine.
XX FT misc_difference 1437..1439
XX FT /*tag= d
XX FT /transl_except= AAG encodes Asparagine.
XX FT misc_difference 1674..1676
XX FT /*tag= e
XX FT /transl_except= CCG encodes Phenylalanine.
XX
XX WO9319090-A.
XX
XX 30-SEP-1993.
XX
XX 16-MAR-1993; 93WO-US02166.
XX
XX 16-MAR-1992; 92GB-0005704.
XX
XX (BARE/) BARENKAMP S J
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Barenkamp SJ;
XX
XX WPI; 1993-320683/40.
XX P-PSDB; AAR41723.
XX
XX High molecular weight surface proteins - of non-typeable
XX haemophilus which exhibit immunogenic properties
XX
XX Claim 3; Figure 1; 100pp; English.
XX
XX The isolation and purification of the high molecular weight protein
XX enables the identification of the major protective epitopes of the
XX protein by conventional epitope mapping. These epitopes can then be
XX synthesised using standard techniques and incorporated into fully
XX synthetic or recombinant vaccines.
XX
XX Sequence 5116 BP; 1766 A; 1008 C; 1037 G; 1305 T; 0 other;
XX
XX Query Match 54.3%; Score 2680.2; DB 14; Length 5116;
XX Best Local Similarity 73.5%; Pred. No. 0;
XX Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps 16;
XX
XX 41 ATGACAAACACATTTACACACCTTTTTCAGCTCTATATGCANATATTTTAAARAAAT 100
XX |
XX 41 ATGACAAACACATTTACACACCTTTTTCAGCTCTATATGCANATATTTTAAARAAAT 99
XX |
XX 101 AGTATAATCCGCATATAAATGGTATATCTTTTCATCTTTTAACTTTTCATC 160
XX |
XX 100 AGTATAATCCGCATATAAATGGTATATCTTTTCATCTTTTAACTTTTCATC 159
XX |
XX 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220
XX |

```

```

160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219
221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 280
220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 279
281 GAATGAAGAGGAGCTGAACGACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 340
280 GAATGAAGAGGAGCTGAACGACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 339
341 TAGGAGAAATATGAACGACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 400
340 TAGGAGAAATATGAACGACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 399
401 TTGCTGTCTGTAATTTGGCAGCGGGTGTGTGACCATTTTCCACAGAAAAGGCTTCCGCTATG 460
400 TTGCTGTCTGTAATTTGGCAGCGGGTGTGTGACCATTTTCCACAGAAAAGGCTTCCGCTATG 459
461 TTACTATCTTTAGGTGTAAACCACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 520
460 CTGCTCGCATGAAAGTGCGTCACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 519
521 TAGGTGTAACATCTATCCCAATCTGTTTTCAGCAAGCGCTTACAGGAATGATGATGATGATGATGAT 580
520 TAGGTGTAACATCTATCCCAATCTGTTTTCAGCAAGCGCTTACAGGAATGATGATGATGATGATGAT 579
581 TACACGGCAGACCCACTATGCAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 640
580 TACACGGCAGACCCACTATGCAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639
641 ACGCTATCATTAATTTGGAAACAAATTTAAACATCGACCAAAATGAAATGATGATGATGATGATGATGAT 700
640 ACGATATCATTAATTTGGAAACAAATTTAAACATCGACCAAAATGAAATGATGATGATGATGATGATGAT 699
701 AAGAAACAAACAACTCCCGCGTATTTTAAACCGGTTTACATCTAACCAAAATGATGATGATGATGATGAT 760
700 AAGAAACAAACAACTCCCGCGTATTTTAAACCGGTTTACATCTAACCAAAATGATGATGATGATGATGAT 759
761 AAGGATTTTAGATTTTAAACGACAAAGTCTTTTAAACCGGTTTACATCTAACCAAAATGATGATGATGAT 820
760 AAGGATTTTAGATTTTAAACGACAAAGTCTTTTAAACCGGTTTACATCTAACCAAAATGATGATGATGAT 819
821 GTAAAGACGCAATTTAAACATCTAACCGGTTTACGCTTACGCTAGACATTTCTAACG 880
820 GTAAAGACGCAATTTAAACATCTAACCGGTTTACGCTTACGCTAGACATTTCTAACG 879
881 AAAAAATCAAGCGCGTATTTTACCTTCGACCAACCAAGATGATGATGATGATGATGATGATGATGATGAT 940
880 AAAAAATCAAGCGCGTATTTTACCTTCGACCAACCAAGATGATGATGATGATGATGATGATGATGATGAT 939
941 TTGTGAATCAGCGTATTTTACCTTCGACCAACCAAGATGATGATGATGATGATGATGATGATGATGATGAT 1000
940 TTGTGAATCAGCGTATTTTACCTTCGACCAACCAAGATGATGATGATGATGATGATGATGATGATGATGAT 999
1001 AAGTGAATAACGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1060
1000 AAGTGAATAACGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1059
1061 AAAAAATCAAGCGGATATTAATAAACCAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1120
1060 AAAAAATCAAGCGGATATTAATAAACCAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119
1121 AAAAAATCAAGCGGATATTAATAAACCAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1180
1120 AAAAAATCAAGCGGATATTAATAAACCAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1179
1181 CTGCGCAATTCGAAACCAAGGTAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1240
1180 CTGCGCAATTCGAAACCAAGGTAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1239
1241 GCAATATTTGTTCTTCCGCAAGAGGGTGAAGCGGAAATTTGCGGCTGTGATGATGATGATGATGATGATGAT 1300
1240 GCAATATTTGTTCTTCCGCAAGAGGGTGAAGCGGAAATTTGCGGCTGTGATGATGATGATGATGATGATGAT 1299

```


QY	1301	AAAAATCAGCAAGCTAAAGCGGCGCAAGCTGATGATTACAGGCGATAAGTGCACATTAAAAA	1361
DB	1300	AAAAAAAAATCAGCAAGCTAAAGCGGCGCAAGCTGATGATTACAGGCGATAAGTGCACATTAAAAA	1359
QY	1361	CAGGTGTCAGTATTACACCTTTTCAGTTAAGAAGGGGGAAGAACTTTACCTTTGGCGGTGACG	1420
DB	1360	CAGGTGTCAGTATTACACCTTTTCAGTTAAGAAGGGGGAAGAACTTTACCTTTGGCGGTGACG	1419
QY	1421	AGCGCGCGCAAGGTTAAAAACGGCATTTCAATTAGCAAGAAAAACCTCTTTAGAAAAAGGCT	1480
DB	1420	AGCGCGCGCAAGGTTAAAAAGGCGATTCAATTTAGCAAGAAAAACCTCTTTAGAAAAAGGCT	1479
QY	1481	CAACCATCAATGTATCAGGCCAAGAAAAAGCGGACGCGCTATTGTGGGCGCATATTG	1540
DB	1480	CAACCATCAATGTATCAGGCCAAGAAAAAGCGGACGCGCTATTGTGGGCGCATATTG	1539
QY	1541	CGTTAATTGACGGCAATTAATAACGCTCAAGCTAGTGGTGCATATCGCTTAAACCGGTGTT	1600
DB	1540	CGTTAATTGACGGCAATTAATAACGCTCAAGGTAGTGGTGATATCGCTTAAACCGGTGTT	1599
QY	1601	TTGTGGAGACATCGGGGCATTTATTTATCCATTGACAGCAATGCAATTTGTTAAAAAAGG	1660
DB	1600	TTGTGGAGACATCGGGGCATTTATTTATCATCAAGACAAATGCAATTTGTGACGCCAAG	1659
QY	1661	AGTGGTTGCTAGACCCCTGATGATGTAAACAATTTGAAGCCGACGACCCCTTGCATATA	1720
DB	1660	AGTGGTTGTTAGACCCGGATATGTATCTATTATGCAAGACAGCAGCAGCGACGACATA	1719
QY	1721	CCGGTATAAATGATGAATTTCCCAACAGCACCGGTGAAGCAAGGACCTTAAAAAATA	1780
DB	1720	CTTCAGAAGACGATGAATACACGGGATCCGGGAATAGTGCCAGCACCCCAACAGCAACA	1779
QY	1781	GCGAACCTCAAAAACACCTTAACCAATACACTATTTTCAAAATATTCGAAAAAGCGCTGGA	1840
DB	1780	AGAGAA--AGACACACTTACAAACACACTCTTGAGAGTATATAAAAAAGGATACCT	1836
QY	1841	CAATGAATATAACGGCATCAAGAAAACTTACCCTTTAATAGCTCAATCAACATCGGAAGCA	1900
DB	1837	TTGTATTACATCACTGCTTAATCAACGCATCTATGTCAATAGCTCCATTATTT--ATCCCA	1893
QY	1901	ACTCCCACTTAATTTCTCCATAGTAAAGTCAAGCTGGCGGAGGCGTTAGATGATGAG	1960
DB	1894	ATGGACCTTAACCTCTTTGGAGCTAGGCTCGGAGCGGTTGGCGGCTTGAGATTAAACAAG	1953
QY	1961	ATATTAC-----TTCTAAAGCGGAAATTAACCATTTATCTGCGGCGATGGG	2008
DB	1954	ATATTACCAACGGTGATGATACAGAGGTGCAACTTAACAATTTACTCAGCGGCGTGGG	2013
QY	2009	TTGATGTTCAATAAATATATTACGTTGATCAGGGTTTTTTAAATATATACCGCGCTTCGG	2068
DB	2014	TTGATGTTTCATAAATAATATCTCACTCGGGCGCAAGGTAAACATAAACATTACGCTAAC	2073
QY	2069	TAGCTTTTGAAGGTGGAAATAACAAAGCAGCGGACGGCAATGCTAAATTTGTCGCC	2128
DB	2074	AGATATCGCCTTTGAAAGAGAGCAACCAAGTC-----ATTACAGGTC	2118
QY	2129	AGGCACCTGTAAACATTACAGGAGAGGAAAAGATTTCAGGGCTAAACACGCTATTCTTAA	2188
DB	2119	AAGGACCTATTACC--TCAGSCAATCAAAAAGGTTTTTAGATTAAATATGCTCTCTAA	2175
QY	2189	ACGGAAACGGTTAAGGTCTGAATATCATTTTCATCAGTCAATAA-----TTTTAACCC	2239
DB	2176	ACGGCACTGGCGACGGACGTGCATAATTCACCACTTAAAGAAACCAATAAATACGCTATCAAA	2235
QY	2240	ACAATCTTTAGTGGCACAATTAAACATATCTGGGAATATAACAAT--TAAACCAAACTACGA	2296
DB	2236	ATAAATTTGAAGGACCTTTTAAATATTTTCAGGGAAGTGAACATCTCAATGGTTTACCTTA	2295
QY	2297	GAAGAACAACCTCGTATGGCCAAACACAGCCATGATTCGCCATCGAACGTCAGTCTCTTA	2356
DB	2296	AAATGAAGGTGATGATATAAATTCAAAGACGACACTTACTCGAATTTAACCTCTCTTAA	2355

QY	2357	ATCT---AGAGACAGCGCGCAATTTTAACTTTTAAATACATTTTCAAGCAANTAGCAAG	2413
Db	2356	ATGTTTCGAGAGTGGCGAGTTTAACCTCACATTTAGCTACTCAGAGGAAGCGATGTCAG	2415
QY	2414	GCTTAACACACAGTATAGAGCTCTGCAGGGGTGAATTTTACGGCGTAAATGGCAACA	2473
Db	2416	GCACACTTACCAGCCTTATATTTAAACGGTATATCATTTCAACAAAGACACT---ACCT	2472
QY	2474	TGTCATTCATCTCAAGAGGAGCGAAAGTTTAATTTCAAATTTAAACCAAGACGAGAACA	2533
Db	2473	TTAATGTTGAACGAAATGCAAGAGTCAACTTTTGACATCAAGGCCACCAATAGGATAATA	2532
QY	2534	TGAACACAGCAACCTTTACCAATTCGGTTTATAGCCAATATCACAGCCACTGTGGGG	2593
Db	2533	AGTATTCAGTTTGAATTAATACGCATCATTTAATGGAACATTTTCAGTTTCGGGAGGGGGA	2592
QY	2594	GCTCTGTTT-----TTTTTGATATATATGCCAACCACTTCCTGCAGAGGGGCTGAGTTAA	2647
Db	2593	GTGTTGATTTCACTTCCTCGCTCATCTCTACGCTCCAAACCCCGGTGTAGTTATAA	2652
QY	2648	AAATCAGTGAATTAATATCTTAACGGCGCTAATTTTACCTTTAAATTTCCCATGTTCBG	2707
Db	2653	ATCTAAATACCTTTAATGTTTCAACAGGCTCAAGTTTAAAGATTTAAACTTCAGGCTCAA	2712
QY	2708	GGGATGAGCGTTTAAAAATCAACAAAGACTTAACCATAAATGCAACCAATTTCAATTTCA	2767
Db	2713	CAAAAACCTGCTCTCAATAGAGANAATTTAACTTTTAATGCCACC-----GAGGCA	2766
QY	2768	GCCTCAGACAGAGAAAGATGATTTTATGACGGGTACGCGCAANTGCCATCAATTCAA	2827
Db	2767	ACATACACATTTTGCAGTTTGAAGCACCAGATGGAATGATTTGTTAAAGCATTTGTAGCCA	2826
QY	2828	CCTACAACATATCCATTTCTGGCGGTAAATGTACCCCTTGTGGACAAACCTCAAGCAGCA	2887
Db	2827	AAAAACATAACCTTTGACGGAGGTAACTACCTTTGGCTCCAGGAAGCGGTAAACG	2886
QY	2888	GCATTCGGGGATTTACTATCGAAGAACGACAAATGTTACGCTAGAAGCCAATAACG	2947
Db	2887	AAATCGAAGCAATGTTACTATCAATTAACACGCTAACGTCACCTCTTATCGGTCGGATT	2946
QY	2948	CCCTTAATCAGCAACATAGGATAGAGTTATAAACTTGGCAGCTTCTCGCTTAATG	3007
Db	2947	TTGACACACATCAAAA-----ACCTTTAACTTTAAAAAGATGTCACTAATAATAGC	3000
QY	3008	GGAGTTTAAGTTTAACTGGCGAAAAATGCGAGATATTAAGGCAATCTCACTATTTCAGAAA	3067
Db	3001	GCAACCTTACCGCTGGAGCAATTTGTCAATATAGCCGGAATCTTACCGTTGAAGTA	3060
QY	3068	GGCCACTTTTAAAGGAAGACTTAGAGATACCTTAATATACCGGCAATTTTACCATA	3127
Db	3061	AGGTAAATTTCAAAGCTATCACAAATTTTCAGTTTAAATGTAAGCGGCTTGTGTGACAACA	3120
QY	3128	ATGGCACTCGCAAAATTAATATAACACAAGAGTGTAAACTTTGGCAATGTTACCAATG	3187
Db	3121	AAAGCAATTCAAATATTTCCATTTGCCAAAGGAGGGCTCGCTTTAAAGACATGTAATTT	3180
QY	3188	ATGTTGATTTAAACATTTACCATTCAGCTTAACGCAACCAAGAACGATCATCGCGGAG	3247
Db	3181	CCAGAATTTTAAGCATCACCACCAACTCCAGCTCCACTTACCGCACTTATTAAGCGGCA	3240
QY	3248	ATATAATCAACAAAAAGGAAGCTTAAATATATACACAGTAAATATGATGCTGAAATCC	3307
Db	3241	ATATAACCAATTAACCGGTGATTTAAATATTTACGAACGAAGTAGTAGATCACTGAAATG	3300
QY	3308	AAATTTGGCGCAATATCTCGCAAAAAGAGCAACCTCAGCATTTCTTCGGTAAAAATTA	3367
Db	3301	AAATTTGGCGCGGATGCTCGCAAAAAAGGAAGTAATCTCACGATTTCTTCTGACAAATCA	3360
QY	3368	ATATACCAACAGATAACCAATCAAAAAGGATTTGATGGAGAGGACTCTAGTTACAGTG	3427
Db	3361	ATATTTACCAACAGATAACCAATCAAGGCAAGCTGTTGTGGGGGAATTCGGATTCAGACG	3420
QY	3428	CGAAGTAATGCCAACCTTAATTTAAACCAAGAAATTTGAAATTTGACAGAAAGCACTAA	3487

Db 3421 CGACAAACAAATGCCAATCTAACCACTTAAACCAAGAAATGAAATTAACGCAAGACTAA 3480
 QY 3488 GTATTTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTAACTATTG 3547
 Db 3481 ATATTTTCAGGTTTCAATTAAGCAGAGATTACAGCTAAGATGGTAGATTAACTATTG 3540
 QY 3548 GCAACAGTAAGTACAGGTCACAGCGTCGCGAAGCCAAACAGTAACTTTTAAACAACTTA 3607
 Db 3541 GTACACCAATAGTCTGA---TGTACTAATGCCAAAAAGTAACCTTTAAACAGGTTA 3597
 QY 3608 AGATTTCAAAATCTCTGCTAGCGTGCACATGTGCACACTAAATAGCAAAAGTGAACACAT 3667
 Db 3598 AGATTTCAAAATCTCTGCTAGCGTGCACATGTGCACACTAAATAGCAAAAGTGAACACAT 3657
 QY 3668 CTAGCAGCAATGGCGACGCTGAAGCAATAGCGGACAGCATACCGGCTTAATCTACTG 3727
 Db 3658 CCGGTAGTAATAACACACTGAAGATAGCAGTGACAATTAATGCCCGCTTAATCTACTG 3717
 QY 3728 CAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAATATCACCGGT 3787
 Db 3718 CAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTGAATCTCTGCGA 3777
 QY 3788 C---GGAAAAGTTTACCACACAGCAGGCTCGACCATTAAGCGCAACAAATGGCAA----- 3839
 Db 3778 CAAGTGGAGAAATTAACCACTAAACAGTACAAACCATTAACGCAACCACTGGTAACGTG 3837
 QY 3840 ----- 3839
 Db 3838 AGATAACCGCTCAACAGGTAGTATCTAGTGGAATTGAGTCCAGCTCTGCTGTGTA 3897
 QY 3840 ----- 3839
 Db 3898 CACTTACTGCAACCGAGGGGCTCTGCTGTAGCAATATTTCGGGCAACACCGTTACTG 3957
 QY 3840 ----- 3839
 Db 3958 TTACTGCAATAGCGGTGCAATTAACCACTTTGGCAGGCTCTACAATTAAGGAACCGAGA 4017
 QY 3840 -AGCAAGTATTACAAACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACACGGTAA 3898
 Db 4018 GTGTAAACCACTTCAAGTCAATCAGCGGTATCGCGGTACGATTTCTGCTGCGCAGCAGTAG 4077
 QY 3899 GTGTAGCGGACTGTGATTAAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATCGG 3958
 Db 4078 AGTTTAAAGCAACCGAAAGTTTAAACCACTCAATCCCAATTAACAAATTAAGCAACACAG 4137
 QY 3959 GTGAGGCTAATGTAAACAGTGCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGG 4018
 Db 4138 CCGAGGCTAACGTAAACAGTGCAACAGGTACAATTTGGGTAGCATTTCCGGTAAATACGG 4197
 QY 4019 TAAATGTACGCAACCGCTGGGATTTACAGTTGGGAATGCGCAGAGAAATTAATGCGA 4078
 Db 4198 TAAATGTACGCAACCGCTGGGATTTACAGTTGGGAATGCGCAGAGAAATTAATGCGA 4257
 QY 4079 CAGAGGAGCTGCAACCTTAACCGCAACAGGTAATCTGACTACTGAGCGGTTCTA 4138
 Db 4258 CAGAGGAGCTGCAACCTTAACCTATCATCTCGGCAATTAATCTACCAAGCTAGTCTAC 4317
 QY 4139 GCATCATTCACTAAGGCTCAGTACGCTTTGGCTTCAGATGGTAGTACCGCAGGAA 4198
 Db 4318 ACATTTACTTTCAGCAAGGCTCAGTAAATCTTCAGCTCAGGATGGTAGGCTTGCAGGAA 4377
 QY 4199 GCATTAATCTGCTAATGTGACATTAATACTACAGGCACCTTAACACCGCTGGCAGGCT 4258
 Db 4378 GTATTAATGCGCCAAATGTGACATAATACTACAGGCACCTTTAACTTACCGTGAAGGGTT 4437
 QY 4259 CGATTAATTAAGCAACCGCAGGCTTGGTATTAAACGCAAAAGATGCTAAAGTAAATG 4318
 Db 4438 CAAACATTAATGCAACCGAGGCTACCTTGGTTATTAAACGCAAAAGACGCTGAGCTAAATG 4497
 QY 4319 GTGATGCACTAGTACAGAGTGAATGCAAGTGAATGCAAGCGGCTCTGGTAGTG 4378

Db 4498 GGCAGAGATTGGTTAAACACACAGCTGGTAATGCAACCAAGCAATGGCTCCGGCAGCG 4557
 QY 4379 TGACTGGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATGGGT 4438
 Db 4558 TAATCGCAGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACCAATTAATGGAT 4617
 QY 4439 TAAATATCATTTTCGAAGATGGTAGAAACACACTGTGCGCTTAAGAGGCAAGAAATTCAGG 4498
 Db 4618 TAAATATCATTTTCGAAGATGGTAGAAACACACTGTGCTGTTAAAAAGGCTTTAAATTTG 4677
 QY 4499 TGAATATATTCAGCGCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACCGCTCC 4558
 Db 4678 TGAATATATTCAGCGCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACCGCTCC 4737
 QY 4559 TTGAAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTAAACCTTGGTGTAA 4618
 Db 4738 TTGAGAAGCTAAAGATTTATCTGATGAAGAAAGAGAAAGCTTACGCTAAACCTTGGAGTAA 4797
 QY 4619 GTGCTGACGTTTGTGAGCCCAATATACAAATTAACATTAACATACACAAATGAATTTA 4678
 Db 4798 GTGCTGACGTTTGTGAGCCCAATATACAAATTAACATTAACATACACAAATGAATTTG 4857
 QY 4679 CAACAGACCGTCAAGTCAAGTGATTAATTTCTGAAGGTAAAGCGGTTTCTCAAGTGTGA 4738
 Db 4858 CAACAGACCGTCAAGTCAAGTGATTAATTTCTGAAGGTAAAGCGGTTTCTCAAGTGTGA 4917
 QY 4739 ATGGCGCAGGATGTACCAATGTGCTGACGATGGACAGCCGTAGTCAGTAAATTTGACA 4798
 Db 4918 ATGGCGCAGGATGTGCGTTAATATCGCTGATTAACGGCGGTAGCGGTAAATTTGACA 4977
 QY 4799 AGTAGATTTCATCCCTGCAATGAAGTCAATTTATTTTCGTATTTATTTCTGTTGGGTTA 4858
 Db 4978 AGTAGATTTCATCCCTGCAATGAAGTCAATTTATTTTCGTATTTATTTCTGTTGGGTTA 5037
 QY 4859 AAGTTCAGTACGGCTTTTACCCATCTGTGAAAAAATTTACGGAGAAATACAAATTAAGTATTT 4918
 Db 5038 AAGTTCAGTACGGCTTTTACCCATCTGTGAAAAAATTTACGGAGAAATACAAATTAAGTATTT 5097
 QY 4919 TTAACAGGTTATTATTATG 4937
 Db 5098 TTAACAGGTTATTATTATG 5116
 RESULT 9
 ID AAQ72293 standard; DNA; 5116 BP.
 AC AAQ72293;
 XX
 XX 25-JUN-1995 (first entry)
 XX
 DE Haemophilus high molecular weight protein HMW1 DNA sequence.
 XX
 KW High molecular weight protein; HMW1; protective vaccine; otitis;
 KW sinusitis; bronchitis; Hib; ss.
 OS Haemophilus.
 XX
 PN W09421290-A.
 XX
 PD 29-SEP-1994.
 XX
 PF 15-MAR-1994; 94WO-US02550.
 XX
 PR 16-MAR-1993; 93US-0038682.
 XX
 PA (BARE/) BARENKAMP S J.
 PA (SGEM/) ST GEME J W.
 XX
 PI Barenkamp SJ, St GEME JW;
 DR WPI; 1994-316665/39.
 DR P-PSDB; AAR63505.

XX

New immunogenic high mol. wt. proteins of non typeable

Hemophilus - useful in protective vaccines

XX

Claim 2; Page 28; 127pp; English.

XX

The Hmwl protein encoded by this sequence is useful in a vaccine to protect against disease caused by non-typeable Haemophilus which are not controlled by H. influenzae type b (Hib) vaccines. The encoded protein can also be used as a carrier for protective Hib polysaccharide (in a conjugate vaccine against meningitis) or for other antigens, haptens, etc. This DNA sequence is part of a larger sequence encoding a gene cluster for the hmw1 gene (9011 bp).

XX

SQ Sequence 5116 BP; 1766 A; 1008 C; 1037 G; 1305 T; 0 other;

Query Match

Best Local Similarity 54.3%; Score 2680.2; DB 15; Length 5116;

Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps

16;

41 ATGACAAACAAATTTACAAACACCTTTTTCAGCTATATGCAAAATATTTAAAAAAT 100

41 ATGACAAACAAATTTACAAACACCTTTTTCAGCTATATGCAAAATATTTAAAAAAT 99

101 AGTATAAATCCGCATATATAAATGGTATATCTTTTCATCTTTCATCTTTCATCTTTC 160

100 AGTATAAATCCGCATATATAAATGGTATATCTTTTCATCTTTCATCTTTCATCTTTC 159

161 TTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 220

160 TTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 219

221 ATCTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 280

220 ATCTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 279

281 GAATGAAGAGGAGCTGAACGAACCAATGATTAAGTAATTTAATTTCACTAACCT 340

280 GAATGAAGAGGAGCTGAACGAACCAATGATTAAGTAATTTAATTTCACTAACCT 339

341 TAGGAGAAAATATGAACAGATATATCTCTCAAAATTCAGAAACGGCTGAATCTTTGG 400

340 TAGGAGAAAATATGAACAGATATATCTCTCAAAATTCAGAAACGGCTGAATCTTTGG 399

401 TTGCTGTCTCTGAATTTGGCAGGGGTTGTGACCATTTCCACAGAAAAGGCTTCGGCTATG 460

400 TTGCTGTCTCTGAATTTGGCAGGGGTTGTGACCATTTCCACAGAAAAGGCTTCGGCTATG 459

461 TTACTATCTTTAGGTGTAAACACCTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 520

460 CTGCTCGCATGAAAGTGGCTCACCTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 519

521 TAGGTGTAAACATCTATTCCCAATCTCTTTAGCAAGCGCTTACAGGAATGGATGTAG 580

520 TAGGTGTAAACATCTATTCCCAATCTCTTTAGCAAGCGCTTACAGGAATGGATGTAG 579

581 TACAGGCACAGCCACTATGCAAGTAGATGGTAAATAAACCAATTTCCGCAACAGTGTG 640

580 TACAGGCACAGCCACTATGCAAGTAGATGGTAAATAAACCAATTTCCGCAACAGTGTG 639

641 ACGTATCATTAATTTGAAACAAATTTAATCATCCACCAATTAATGGTGGCAGTTTTTAC 700

640 ACGTATCATTAATTTGAAACAAATTTAATCATCCCAATTAATGGTGGCAGTTTTTAC 699

701 AAGAAAACAACTCCGCGTATTTCAACCGTGTTCATCTCAACCAATTTCCCAATTA 760

700 AAGAAAACAACTCCGCGTATTTCAACCGTGTTCATCTCAACCAATTTCCCAATTA 759

761 AAGGATTTTAGATTTCAACGGACAGTCTTTTAAATCAACCCAAATGGTATCACAATAG 820

760 AAGGATTTTAGATTTCAACGGACAGTCTTTTAAATCAACCCAAATGGTATCACAATAG 819

821 GTAAAGACGCAATTTAATACACTAATGGCTTTACGGCTTCTACGGTAGACATTTCTAACG 880

820 GTAAAGACGCAATTTAATACACTAATGGCTTTACGGCTTCTACGGTAGACATTTCTAACG 879

881 AAAACATCAAGCGCGCTAATTTTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAA 940

880 AAAACATCAAGCGCGCTAATTTTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAA 939

941 TTGTGAATCAGCGTTAATTTACTGTGCGTAAGAGCGGAGTGTAAATCTTTATTGGTGSCA 1000

940 TTGTGAATCAGCGTTAATTTACTGTGCGTAAGAGCGGAGTGTAAATCTTTATTGGTGSCA 999

1001 AAGTGAATAACGAGGCTGTGATTAGCGTAAATGGTGGCAGCATTTCTTTACTCCAGGGC 1060

1000 AAGTGAATAACGAGGCTGTGATTAGCGTAAATGGTGGCAGCATTTCTTTACTCCAGGGC 1059

1061 AAAAATCACCATCAGCGATATAATAAACCCCAACCATTTACTTACAGCATTTGCCCGCGCTG 1120

1060 AAAAATCACCATCAGCGATATAATAAACCCCAACCATTTACTTACAGCATTTGCCCGCGCTG 1119

1121 AAAATGAAGCGGTCAATCTGGCGCATATTTTGGCCAAAGCGGTTAAATTTGTCGGTG 1180

1120 AAAATGAAGCGGTCAATCTGGCGCATATTTTGGCCAAAGCGGTTAAATTTGTCGGTG 1179

1181 CTGCCACTATTCGAAACCAAGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAGCG 1240

1180 CTGCCACTATTCGAAACCAAGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAGCG 1239

1241 GCAATATTTGTTCTTTCGCCCAAGAGGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1300

1240 GCAATATTTGTTCTTTCGCCCAAGAGGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1299

1301 AAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTACAGCGCATAAAGTCACATTAATAA 1360

1300 AAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTACAGCGCATAAAGTCACATTAATAA 1359

1361 CAGGTGCAAGTTTTCGACCTTTTCAAGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1420

1360 CAGGTGCAAGTTTTCGACCTTTTCAAGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1419

1421 AGCGCGCGAAGGTAAACCGGCAATTTCAATTAGCAAGCAAAACCTCTTTAGAAAGGCT 1480

1420 AGCGCGCGAAGGTAAACCGGCAATTTCAATTAGCAAGCAAAACCTCTTTAGAAAGGCT 1479

1481 CAACCATCAATGTATCAGGCAAAAGAAAGCGGACCGCTATTGTTGGGCGGATATTG 1540

1480 CAACCATCAATGTATCAGGCAAAAGAAAGCGGACCGCTATTGTTGGGCGGATATTG 1539

1541 CGTTAATTCAGCGCAATATTACCGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTT 1600

1540 CGTTAATTCAGCGCAATATTACCGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTT 1599

1601 TTGTGGAGACATCGGGGCAATTTATTTATCCATTTGACAGCAATCAATTTGTTAAACCAAG 1660

1600 TTGTGGAGACATCGGGGCAATTTATTTATCCATTTGACAGCAATCAATTTGTTAAACCAAG 1659

1661 AGTGGTTGCTAGACCTGTATGATGTAAACAAATTTGAAGCGGAGACCCCTTCGCAATAATA 1720

1660 AGTGGTTGCTAGACCTGTATGATGTAAACAAATTTGAAGCGGAGACCCCTTCGCAATAATA 1719

1721 CCGGTATAAATGATGAATTTCCCAACAGCGCACCGGTGAAGCAGCGACCTTAAAAAATA 1780

1720 CTTTCAAGACCATGATATACAGCGGATTCGGGAATAGTGGCAGCACCCCAACCAAGCA 1779

1781 GCGAATCAAAACAGCTTAACCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1840

1780 AAGAA---AAGACAACATTAACAAACCAACTCTTTGAGAGTATCTTAAAGAGTACT 1836

1841 CAATGAATATACGGCATCAAGAAACCTTACGGTTAATAGCTCAATCAACATCGGAAGCA 1900

1837 TTGTTAATCAGCTCAATCAACCGCATCTATGTAATAGCTCAATTTAATTT---ATCCA 1893

1901 ACTCCCACTTAATTTCTCCATAGTAAAGGTACAGGTGGCGGAGCGTTCAGATTCATGAG 1960

Db 1894 ATGGCAGCTTAACCTCTTTGGAGTGGGGTGGAGGGTGGGGCGGTGAGATTAACAACG 1953
Qy 1961 ATATTAC-----TTCTAAAGCGGAAATTTAAACCATTTATCTGCGCGATGGG 2008
Db 1954 ATATTACCGCGGTGATGATACAGAGGTGCAAACTTAACAATTTACTCAGCGCGCTGGG 2013
Qy 2009 TTGATGTTTCAATAAATAATTTACGCTTGATCAGGGTTTTTTTAAATATTTACCGCGCTCGG 2068
Db 2014 TTGATGTTTCAATAAATAATTTACTCTCGGGCGCAAGGTAAACATTAACAGCTAAAC 2073
Qy 2069 TAGCTTTTGAAGTGGAAATAACAAGACGCGAGCGGCAAAATGCTAAATTTGTCGCC 2128
Db 2074 AAGATATCGCTTTGAGAAAGGAACAACCAAGTC-----ATTACAGTTC 2118
Qy 2129 AGGCACTGTAAACATTTACAGGAGAGGGAAGATTTTCAGGGCTCAACAACGTATCTTTAA 2188
Db 2119 AAGGCTATTAC---TCAGGCAATCAAAAGGTTTAGATTTAATATGCTCTCTAA 2175
Qy 2189 ACGGAACGGGTAAAGTCTGTAATATCAATTTTCATCAGTGAATAA-----TTTAAACC 2239
Db 2176 ACGCACTGGCAGCGGACTGCAATTCACCACATAAGAACCAATAAATAGCTATCAAA 2235
Qy 2240 ACAATCTTAGTGGCAAAATTAACATATCTGGGAATATAACAAT---TAACCAAACTACGA 2296
Db 2236 ATAAATTTGAAGGACCTTTAAATATTTTCAGGGAAGTGAACATCTCAATGGTTTTACCTA 2295
Qy 2297 GAAAGAACACCTCTGATTTGGCAACACAGCCATGATTCGCACCTGGAAGTCAAGTCTTA 2356
Db 2296 AAAATGAAAGTGGATATAATAATTTCAAGGAGCGCACTTACTGGAATTTAAACCTCTTAA 2355
Qy 2357 ATCT---AGACAGCGCGCAAAATTTTACCTTTATTAATATACATTTCAAGCAATAGCAAG 2413
Db 2356 ATGTTTCCGAGAGTGGGAGTTTAACTCTACTATTGACTCCAGAGGAAGGATAGTGCAG 2415
Qy 2414 GCTTAACAACACAGTATAGAGCTCTCAGGGGTGAATTTTAAAGCGGTAAATTTAAAGCGCA 2473
Db 2416 GCACACTTACCGCTTTAATTTAAACGGTATATCATTTCAACAAAAGACACT---ACCT 2472
Qy 2474 TGTCATTAATCTCAAGAGGAGGGAAGTTAATTTCAATTTAAACCAACAGAGAA 2533
Db 2473 TTAATGTTGAAGCAATGCAAGAGTCAACTTTTGACATCAAGGCACCAATAGGGATAAATA 2532
Qy 2534 TGAACAAGCAACCTTTACCAATTCGGTTTTAGCCAAATATCACGCCACTGTCGGG 2593
Db 2533 AGTATCTAGTTTGAATTCAGCATCAATTAATGGAACATTTTCAGTTTCGGGAGGGGGA 2592
Qy 2594 GCTCTGTTT-----TTTTTGATATATATGCCAACCTTCTGGCAGAGGGCTGAGTTAA 2647
Db 2593 GTGTTGATTTACACTTCTCGCCTCATCTTAAGCTCCAAACCCCGGTGTAGTTATAA 2652
Qy 2648 AAATGAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTTAAATTTCCCATGTTCCGG 2707
Db 2653 ATTTCAATAATCTTTAATGTTTCAACAGGGTCAAGTTTAAAGATTTAAACCTTTACGGCTCAA 2712
Qy 2708 GCGATGACGCTTTTAAATCAACAAGACTTAACCATAAATGCAACCAATTTCAAAATTTCA 2767
Db 2713 CAAAAACTGGCTTCTCAATAGAGAAGATTTAACTTTAAATGCCACC-----GGAGGCA 2766
Qy 2768 GCTTCAGACAGAGAAAGATGATTTTTTATGACGGGTACGCGCAATGCCATCAATTTCAA 2827
Db 2767 ACATAACACTTTTGAAGTTGAAGCCAGCGATGGAATGATTTGTTAAAGGCATTTGAGCCA 2826
Qy 2828 CCTACACATATCACTTGGCGGTAAATGTCACCTTTGGTGACAAACCAACAGCAGCA 2887
Db 2827 AAAAACAATAACCTTTGAGGAGGTACATCACTTTGGTCCAGGAAGCCGTAAACAG 2886
Qy 2888 GCATTACGGGGAAATTAATCTATGAGAAAGACGAAATTTAGCGTAAAGCCATAACG 2947
Db 2887 AAATCGAAGCAATGTTACTATCAATAAACAACGCTAACGCTCACTCTTATCGGTTCCGAT 2946
Qy 2948 CCCCTAATCAGCAAAACATAAGGGATAGAGTTTATAAACTTTGGCAGCTTGGCTGTTAATG 3007
Db 2947 TTGACAACCATCAAAA-----ACCTTTAACTATTAAAAAAGATGTCATCAATTAATAGCG 3000

Qy 3008 GGAGTTAAGTTTAACTGGCGAAATGCAGATATTAAGGCAATCTCACATATTTTCAGAAA 3067
Db 3001 GCACCTTTCGCGTGGAGCAATATTGTCAATATAGCGGAAATCTTACCGTTGAAAGTA 3060
Qy 3068 GCGCACCTTTTAAAGGAAGACTAGAGTACCCTAAATATCACCGCAATTTTACCATAA 3127
Db 3061 ACGCTAAATTTCAAGCTATCAAAATTTTCATTTTAAATGAGCGGCTGTTTGACAACA 3120
Qy 3128 ATGGCAGTCGCGAAATTAATATACACAGAGGTGGTAAAACTTTGGCAATCTTACCAGT 3187
Db 3121 AAGGCAATTTCAAAATTTCCCATTTGCCAAAGGAGGGGCTCGCTTTAAAGCAATTAAT 3180
Qy 3188 ATGGTGATTTAAACATTACCACCTCACGCTAAACGCAACCAAGAGCATCATCGGCGAG 3247
Db 3181 CCAAGAAATTTAAGCATCACCACTCCAGCTCCACCTTACCGCACTATTATAGCGGCA 3240
Qy 3248 ATATAATCAACAAAAAAGAGCTTTAAATATTATACAGACAGTAATAATGATGCTGAAATCC 3307
Db 3241 ATATAACCAATAAAAACGGTGATTTTAAATATTACGAAGAGGTAGTATGATGTAATATG 3300
Qy 3308 AAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTCACGATTTCTTCCGATAAAATTA 3367
Db 3301 AAATTTGGCGCGCTGTCGCAAAAAGAGGTAATCTCACGATTTCTTGTGACAAAATCA 3360
Qy 3368 ATATCACCAACAGATTAACAATCAAAAGGGTATTGTAGGAGGACTCTAGTTCAGATG 3427
Db 3361 ATATACCAACAGATTAACAATCAAAAGGGTATTGTAGGAGGAGTATTCGGGAGAAAT 3420
Qy 3428 CGACAAGTAAATGCGCAACCTTAATTTAAACCAAGAAATTAAGATTTGACAGAGACCTAA 3487
Db 3421 CGACAACCAATGCGCAATTAATTTAAACCAAGAAATTAAGATTTGACAGAGACCTAA 3480
Qy 3488 GTATTTAGTGTTCATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTAACTATTG 3547
Db 3481 ATATTTACAGTTTCAATAAAGCAGAGATTACAGTCAAGGTGATGATGATTTAACTATTG 3540
Qy 3548 GCAACAGTAAATGAGCGTAAACAGCGGTGCGCAAGGCAACAGTAACTTTTAACTATTG 3607
Db 3541 GTAACACCAATAGTGTCTGA---TGGTACTTAATGCCAAAAGAAATTTTAACTATTG 3597
Qy 3608 AAGATTCAAAAATCTCTGCTGACGGTCAACAATGTGACACTTAAATAGCAAGTAAACAT 3667
Db 3598 AAGATTCAAAAATCTCTGCTGACGGTCAACAAGGTGACACTTACACAGCAAGTAAACAT 3657
Qy 3668 CTAGCAGCAATGGCGGACGTGAAAGCAATAGCGACACAGTACCGGCTTAACTATTACTG 3727
Db 3658 CCGGTAGTAATAACAACACTGAAGATAGCATGACAATAATGCGGCTTAACTATTGATG 3717
Qy 3728 CAAAAATGTAGAGTAACAACAAGATATTACTTCTCAAAAACAGTAAATATCACCGCGT 3787
Db 3718 CAAAAATGTAGAGTAACAACAAGATATTACTTCTCAAAAACAGTAAATATCACCGCGT 3777
Qy 3788 C---GGAAAAGGTACCAACAGCGCTCGACCTTAAAGCAACCAATTTGCAAAAGTAA 3839
Db 3778 CAAGTGAGAAATTTACCCCTTAACACAGTACACCACTTACGCAACCACTGGTAACTGG 3837
Qy 3840 ----- 3839
Db 3838 AGATAACCGCTCAAAACAGGTAGTATCTAGTGGAAATTTAGTCCAGCTCTGGCTCTATA 3897
Qy 3840 ----- 3839
Db 3898 CACTTACTGCAACCGAGGGGCGCTCTTGTGTAGCAATAATTTTCGGGCAACACCGTTACTG 3957
Qy 3840 ----- 3839
Db 3958 TTACTGCAATAGCGGTGCTAATAACCACTTTTGGCAGGCTCTACAATTTAAAGAACCGAGA 4017
Qy 3840 -AGCAAGTATTACAAACCAAAACAGTATATCACGGTACAGATTTCCGGTTAAACAGGTAA 3898
Db 4018 GTGTAACCACTTCAAGTCAANTCAGCGCATATCGCGGTACAGATTTCTTGGTGGCACAGTAG 4077

```
QY 3899 GTGTTAGCGGACTGGTGATTTTAAACACATAAATCCGGCTCAAAAATGAAGCGAAATCGG 3958
Db 4078 AGGTTAAAGCAACCAAGTTTAAACCACTCAATCCAAATTAAGCAACAACAG 4137
QY 3959 GTGAGGCTAATGTAAACAAGTGAACAGGTACAAATGGCGGTACAAATTCGCGTAATACGG 4018
Db 4138 GCGAGGCTAACGTAAACAAGTGAACAGGTACAAATGGCGGTACAAATTCGCGTAATACGG 4197
QY 4019 TAAATGTTACGGCAACCGTGGCGATTTAACAGATTGGGAATGGCGAGAAATTAATGCGA 4078
Db 4198 TAAATGTTACGGCAACCGTGGCGATTTAACAGATTGGGAATGGCGAGAAATTAATGCGA 4257
QY 4079 CAGAAGGAGCTGCAACCTTTAACCGCAACAGGAAATACCTTGACTACTGAAGCGGTTCTA 4138
Db 4258 CAGAAGGAGCTGCAACCTTTAACCGCAACAGGAAATACCTTGACTACTGAAGCGGTTCTA 4317
QY 4139 GCATCACTTCAACTAAGGCTAGGTAGACCTCTTGCGCTCAGAAATGGTAGCATCGCAGGAA 4198
Db 4318 ACATTAATCTCAGCAAGGCTAGGTAAATCTTTGAGCTCAGGATGGTAGCGTTCGAGGAA 4377
QY 4199 GCATTAATCTGCTAATGTGACATTTAAATACCTACAGGCACTTAAACCACTGGCAGGCT 4258
Db 4378 GTATTATGCTCCGCAATGTGACATTTAAATACCTACAGGCACTTAAACCACTGGCAGGCT 4437
QY 4259 CGGATATTAAAGCAACCGTGGCGATTTAACAGATTGGGAATGGCGAGAAATTAATGCGA 4318
Db 4438 CAACATTAATGCAACCGTGGCGATTTAACAGATTGGGAATGGCGAGAAATTAATGCGA 4497
QY 4319 GTGATGATCAGTGTATAGTACAGAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4378
Db 4498 GCGCAGCATTTGGGTAAACCAACAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4557
QY 4379 TGACTGCGGCAACCTCAACAGGCTGGAATATCACTGGGATTTAAACACAGTAAATGGT 4438
Db 4558 TAATCGCACAACCTCAACAGGCTGGAATATCACTGGGATTTAAACACAGTAAATGGT 4617
QY 4439 TAAATATCATTTGCAAGATGGTGAACACACTGCTGCGTTAAAGAGGCAAGAAATTCAGG 4498
Db 4618 TAAATATCATTTGCAAGATGGTGAACACACTGCTGCGTTAAAGAGGCAAGAAATTCAGG 4677
QY 4499 TGAATATATCCAGCAGGCTGACAGTGTAGAAAGTGAATGAAGCGAAACGCGTCC 4558
Db 4678 TGAATATATCCAGCAGGCTGACAGTGTAGAAAGTGAATGAAGCGAAACGCGTCC 4737
QY 4559 TTGAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAACTGGGTAA 4618
Db 4738 TTGAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAACTGGGTAA 4797
QY 4619 GTGCTGTACGTTTGTGAGCCAAATTAATCAATTCAGTCAATACACAAATCAATTTA 4678
Db 4798 GTGCTGTACGTTTGTGAGCCAAATTAATCAATTCAGTCAATACACAAATCAATTTA 4857
QY 4679 CAACGAGACCGTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTCTTCTCAAGTGGTA 4738
Db 4858 CAACGAGACCGTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTCTTCTCAAGTGGTA 4917
QY 4739 ATGCGGACAGGATGTACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACA 4798
Db 4918 ATGCGGACAGGATGTACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACA 4977
QY 4799 AGGTAGATTTTCCTGCAATGAAGTCAATTTTATTTCTGATTTATTTACTGTGGGTGA 4858
Db 4978 AGGTAGATTTTCCTGCAATGAAGTCAATTTTATTTCTGATTTATTTACTGTGGGTGA 5037
QY 4859 AAGTTAGTACGGGCTTTACCCATCTTGTAAAAAATACGGAGAAATACAAATAAAGTATTT 4918
Db 5038 AAGTTAGTACGGGCTTTACCCATCTTGTAAAAAATACGGAGAAATACAAATAAAGTATTT 5097
QY 4919 TTAACAGGTTATTATATG 4937
Db 5098 TTAACAGGTTATTATATG 5116
```

```
RESULT 10
AAT90994
ID AAT90994 standard; DNA; 5116 BP.
XX AC AAT90994;
XX DT 14-APR-1998 (first entry)
XX DE Non-typeable Haemophilus high mol.wt. surface protein hmw1 gene.
XX KW Non-typeable Haemophilus; high molecular weight surface protein;
XX KW hmw1 gene; immunogen; vaccine; otitis media; ss.
XX OS Haemophilus influenzae strain 12.
XX PH
XX Key Location/Qualifiers
RBS 341..345
FT /*tag= a
FT 129..135
FT repeat_unit
FT /*tag= b
FT rpt_type= TANDEM
FT 136..142
FT /*tag= c
FT rpt_type= TANDEM
FT 143..149
FT /*tag= d
FT rpt_type= TANDEM
FT repeat_unit
FT 150..156
FT /*tag= e
FT rpt_type= TANDEM
FT 157..163
FT /*tag= f
FT rpt_type= TANDEM
FT repeat_unit
FT 164..170
FT /*tag= g
FT rpt_type= TANDEM
FT 171..177
FT /*tag= h
FT rpt_type= TANDEM
FT repeat_unit
FT 178..184
FT /*tag= i
FT rpt_type= TANDEM
FT 185..191
FT /*tag= j
FT rpt_type= TANDEM
FT repeat_unit
FT 192..198
FT /*tag= k
FT rpt_type= TANDEM
FT repeat_unit
FT 199..205
FT /*tag= l
FT rpt_type= TANDEM
FT 206..212
FT /*tag= m
FT rpt_type= TANDEM
FT repeat_unit
FT 213..219
FT /*tag= n
FT rpt_type= TANDEM
FT repeat_unit
FT 220..226
FT /*tag= o
FT rpt_type= TANDEM
FT 227..233
FT /*tag= p
FT rpt_type= TANDEM
FT repeat_unit
FT 234..240
FT /*tag= q
FT rpt_type= TANDEM
FT CDS
FT 351..4961
FT /*tag= r
FT /*transl_except= (pos:360..362, aa:Ile)
FT /*transl_except= (pos:642..644, aa:Ala)
FT /*transl_except= (pos:1437..1439, aa:Asn)
FT repeat_unit
FT /*tag= s
```

```

XX      /rpt_type= INVERTED
XX      W09736914-AL.
XX      09-OCT-1997.
XX      01-APR-1997; 97WO-US04707.
XX      01-APR-1996; 96US-0617697.
XX      (BARE/) BARENKAMP S J.
XX      Barenkamp SJ;
XX      WPI: 1997-503038/46.
XX      P-PSDB; AAW30293.
XX      High molecular weight proteins of non-typeable Haemophilus
XX      influenzae - useful for vaccine production
XX      Claim 6; Page 63-66; 183pp; English.
XX      This nucleic acid comprises the hmw1 gene of non-typeable
XX      Haemophilus influenzae strain 12 that encodes high molecular
XX      weight surface protein HMW1 (see AAW30293). A phage genomic
XX      library of strain 12 was screened for clones expressing high
XX      mol.wt. proteins using a high titre antiserum against HMWs.
XX      Strongly reactive clones were subcloned into T7 expression
XX      plasmid; all expressed either 125 kDa HMW1 or 120 kDa HMW2 (see
XX      AAW30294). The expressed proteins are truncated, starting at
XX      residue 442 of both full-length HMW1 and HMW2 gene products.
XX      Correct processing requires the products of additional
XX      downstream genes (see AAT90996 and AAT90997). Nucleotide sequences
XX      (see AAT90992-93) encoding 2 HMW proteins of non-typeable H.
XX      influenzae strain 5 (see AAW30291-92) have also been identified.
XX      The HMW proteins, conjugates and peptides can be used in
XX      vaccines, as immunogens for preparation of antibodies and as
XX      antigens for detection of these antibodies. The nucleic acid
XX      sequences can be used as to prepare recombinant proteins and as
XX      probes for detection of related genes.
XX      Sequence 5116 BP; 1766 A; 1009 C; 1036 G; 1305 T; 0 other;
XX      Query Match 54.38; Score 2680.2; DB 18; Length 5116;
XX      Best Local Similarity 73.58; Pred. No. 0;
XX      Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps
XX      16;
XX      41 ATGACAAACACAAATACAAACACCTTTTTCAGTCTATATGCAAAATATTTAAAAAAT 100
XX      41 ATGACAAACACAAATACAAACACCTTTTTCAGTCTATATGCAAAATATTT-AAAAAT 99
XX      101 AGTATAATCCGCCATATAAATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTCATC 160
XX      100 AGTATAATCCGCCATATAAATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTCATC 159
XX      161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTC 220
XX      160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTC 219
XX      221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTT 280
XX      220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTT 279
XX      281 GAATCAAGAGGAGCTGAACGACCAATATGATTAAGTAAATTTAATTTCTTCAACTTAACCT 340
XX      280 GAATCAAGAGGAGCTGAACGACCAATATGATTAAGTAAATTTAATTTCTTCAACTTAACCT 339
XX      341 TAGGAGAAATATGACAGATATATCTTCAATTCAGCAAGAGGAGGAGGAGGAGGAGGAGGAG 400
XX      340 TAGGAGAAATATGACAGCTATATCTTCAATTCAGCAAGAGGAGGAGGAGGAGGAGGAGGAG 399
XX      401 TTGCTGTGTCGAATTTGGCAGCGGGTGTGACCATTTCCACAGAAAGAGGCTTCGCGTATG 460
XX      400 TTGCTGTGTCGAATTTGGCAGCGGGTGTGACCATTTCCACAGAAAGAGGAGGAGGAGGAGGAG 459
XX      461 TTACTATCTTTAGGTGTAACCACTTAGCGTTAAAGCCCACTTTCCGCTATGTTACTATCTT 520
XX      460 CTGCTCGCATGAAAGTGGTGCATTTAGCGTTAAAGCCCACTTTCCGCTATGTTACTATCTT 519
XX      521 TAGGTGTAACATCTATTCCACAATCTGTTTACGAGCGGCTTACAGCAAGTGGATGATAG 580
XX      520 TAGGTGTAACATCTATTCCACAATCTGTTTACGAGCGGCTTACAGCAAGTGGATGATAG 579
XX      581 TACAGGCGACAGCCACTATGCAAGTAGATGTAATAAACCATTATTCGCAACAGCTGTTG 640
XX      580 TACAGGCGACAGCCACTATGCAAGTAGATGTAATAAACCATTATTCGCAACAGCTGTTG 639
XX      641 AGCTATCATTAATTGGAAACCAATTTAACATCGACCAAAATGAAATGATGTCAGTTTTTAC 700
XX      640 AGCATATCATTAATTGGAAACCAATTTAACATCGACCAAAATGAAATGATGTCAGTTTTTAC 699
XX      701 AAGAAACCAACCAACTCCGCGCTATTCAACCGTGTATCACTTAAACCAAAATCTCCCAATTA 760
XX      700 AAGAAACCAACCAACTCCGCGCTATTCAACCGTGTATCACTTAAACCAAAATCTCCCAATTA 759
XX      761 AAGGGATTTTAGATTTCTAACGACAGTCTTTTAAATCAACCAAAATGATGTCACAATAG 820
XX      760 AAGGGATTTTAGATTTCTAACGACAGTCTTTTAAATCAACCAAAATGATGTCACAATAG 819
XX      821 GTAAAGACCAATTTAACTACACTTAATGCGCTTTACGCTTCTAGCTAGACATTTCTTAAGC 880
XX      820 GTAAAGACCAATTTAACTACACTTAATGCGCTTTACGCTTCTAGCTAGACATTTCTTAAGC 879
XX      881 AAACATCAAGGCGCTAATTTACCTTCGACCAACCAAAAGATGATGTCGCTGCTGCTGCTG 940
XX      880 AAACATCAAGGCGCTAATTTACCTTCGACCAACCAAAAGATGATGTCGCTGCTGCTGCTG 939
XX      941 TTCTGAATCACGGTTTAAATCTGTCGTAAGACGCGGAGTGTAAATCTTTATTGTTGGGCA 1000
XX      940 TTCTGAATCACGGTTTAAATCTGTCGTAAGACGCGGAGTGTAAATCTTTATTGTTGGGCA 999
XX      1001 AAGTGAACACGAGGCTGTGATTAGCTAAATGTTGGGAGCATTTCTTTACTGCGAGGCG 1060
XX      1000 AAGTGAACACGAGGCTGTGATTAGCTAAATGTTGGGAGCATTTCTTTACTGCGAGGCG 1059
XX      1061 AAAAAATCACCATCACGATATATAAACCACCAACCATTTACTTACAGCATTTGCGCGCGCTG 1120
XX      1060 AAAAAATCACCATCACGATATATAAACCACCAACCATTTACTTACAGCATTTGCGCGCGCTG 1119
XX      1121 AAAATGAAGCGGTCAATCTGGCGGATATTTTTCGCAAGCGGTAAACATTAATGTCGCTG 1180
XX      1120 AAAATGAAGCGGTCAATCTGGCGGATATTTTTCGCAAGCGGTAAACATTAATGTCGCTG 1179
XX      1181 CTGCCACTATTCGAAACCAAGGTAAACTTCTGCTGATCTGTAAGCAAGATGAAAGCG 1240
XX      1180 CTGCCACTATTCGAAACCAAGGTAAACTTCTGCTGATCTGTAAGCAAGATGAAAGCG 1239
XX      1241 GCAATATGTTCTTCGCGCAAGAGGTGAAGGAAATTTGGGCTGTAATTTCCGCTC 1300
XX      1240 GCAATATGTTCTTCGCGCAAGAGGTGAAGGAAATTTGGGCTGTAATTTCCGCTC 1299
XX      1301 AAAATCAGCAAGCTTAAAGCGGCAAGCTGATGATTAACGCGGATPAAAGTCAATTAATAA 1360
XX      1300 AAAATCAGCAAGCTTAAAGCGGCAAGCTGATGATTAACGCGGATPAAAGTCAATTAATAA 1359
XX      1361 CAGGTGCAAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGGCGGTGACG 1420
XX      1360 CAGGTGCAAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGGCGGTGACG 1419
XX      1421 AGCGCGCGAAGGTAAAGCGGCAAGCTGATGATTAACGCGGATPAAAGTCAATTAATAA 1480
XX      1420 AGCGCGCGAAGGTAAAGCGGCAAGCTGATGATTAACGCGGATPAAAGTCAATTAATAA 1479
XX      1481 CAACCATCAATGATCAGGCAAGCAAGCAAGCGGAGCGCTATTGTTGGGCGGATATTG 1540
XX      1480 CAACCATCAATGATCAGGCAAGCAAGCAAGCGGAGCGCTATTGTTGGGCGGATATTG 1539

```

Qy	1541	CGTTAAATTGACGCGCAATATTAAACGCTCAAGCGTAGTGGTATATCGCTAAAAACCGGTGTT	1600
Db	1540	CGTTAAATTGACGCGCAATATTAAACGCTCAAGGTAGTGGTATATCGCTAAAAACCGGTGTT	1599
Qy	1601	TTGTGGAGACATCGGGCATATTATTCATTTGACAGCAATGCAATTTAAAAACAAAG	1660
Db	1600	TTGTGGAGACATCGGGGCATGATTATTCATCAAAAGACAATGCAATTTGACGCCAAAG	1659
Qy	1661	AGTGGTTGCTAGACCCCTGATGTAAACAATTTGAAGCCGAGACCCCTTCGCAATATA	1720
Db	1660	AGTGGTTGCTAGACCCGGATATGTATCTATTATGCAAGACAGCAGCAGCAGCAATA	1719
Qy	1721	CCGGTATAAATGATGAATTTCCCAACAGGCACCGGTGAAGCAAGCCACCTTAAAAAATA	1780
Db	1720	CTTCAGAAGACGATGAATACAGGGATCCGGGAATAGTCCAGCACCCCAAAACGAAACA	1779
Qy	1781	GCGAACCTCAAAACAACGCTTAACCAATACAACTATTTCAAATTATCTGAAAACCCCTTGA	1840
Db	1780	AGAA---AGACACACATTACAAACACACCTCTTTGAGAGTATACTAAAAAAGGTACCT	1836
Qy	1841	CAATGAATATAAGCGCATCAAGAAACTTTACCGTTAATAGCTCAATCAACATCGGAAGCA	1900
Db	1837	TTGTTAACATCACTGCTAATCAACGCATCTATGTCAATAGCTCCATTAATTT---ATCCA	1893
Qy	1901	ACTCCACACTTAATTTCTCCATAGTAAAGGTCACGCTGCGGAGGCGTTCAGATTGATGGAG	1960
Db	1894	ATGCGACCTTAACCTCTTTGAGGTGAGGCTCGGCGGTGGCGCGTTGAAGATTAAACAAG	1953
Qy	1961	ATATTAC-----TTCTAAAGCGGAATTTAAACCATTTATCTTGGCGGATGGG	2008
Db	1954	ATATTACACCGGTGATGATACAGAGGTGCAAACTTAACAATTTACTCAGCGCGCTGG	2013
Qy	2009	TTCATGTTTCATAAAATATTACGCTTGATCAGGGTTTTTTAAATATTACCGCGCTTCGG	2068
Db	2014	TTCATGTTTCATAAAATATCTCCTCGGGCGCAAGGTAACATAAACATTTACAGCTAAAC	2073
Qy	2069	TAGCTTTTGAAGGTGGAATATAACAAGCACGCGACGGCGAAATGCTTAAATTTGTCGCC	2128
Db	2074	AAGATATCGCTTTTGAGAAAGGAACCAACCAAGTC-----ATTACAGGTC	2118
Qy	2129	AGGCACTCTAACCATTTACAGGAGAGGAAAAGATTTTCAGGGCTAACACGTAATCTTAA	2188
Db	2119	AAGGCACTATTACC---TCAGCAATCAAAAAGGTTTTAGATTTAATATGCTCTCTAA	2175
Qy	2189	ACGGAACGGGTAAAGGTCTGAATATCAATTTTCATCAGTGAATA-----TTTAAACC	2239
Db	2176	ACGGCACTGGCAGCGACTGCAATTTACCACATAAAGAACAATAATAGCTATCACAA	2235
Qy	2240	ACAATCTTAGTGGCACAATTACATACTGGGAATATAACAAT---TAACCAACTACGA	2296
Db	2236	ATAAATTTGAAGGGACTTTAAATATTTCAGGGAAGTGAACATCTCAATGGTTTTACCTA	2295
Qy	2297	GAAAGAACACCTCTGATTTGGCAAAACACGCCATGATTCGCACCTGGAAGCTCAGTGCTTA	2356
Db	2296	AAAATGAAAGTGGATATGATAATAATTCAAAGGACGCCTTACTGGAATTTAACCTCTTAA	2355
Qy	2357	ATCT---AGACACAGGCCAAATTTTACCTTTATTAATACATTTCAAGCAATAGCAAG	2413
Db	2356	ATGTTTCGAGAGTGGCGAGTTTAACTCTACTATTTGACTCCAGGAAGCGATGTGCAG	2415
Qy	2414	GCCTTAACACACAGTATAGAAGCTCTGCAGGGTGAATTTTAAACGCGTAAATGGCAACA	2473
Db	2416	GCACACTTACCAGCCTTATTAATTTAAACGGTATATCATTTCAACAAGACACT---ACCT	2472
Qy	2474	TGTCATTCAAATCTCAAGAAGGACGGAAGTTAATTTCAAATTTAAAAACCAACGAGAACA	2533
Db	2473	TTAATTTGTAACGAAATGCAAGAGTCAACTTTTGACATCAAGGCCAATATAGGATATAATA	2532
Qy	2534	TGAACACACAGCAACCTTTACCATTCCGTTTTTACGCAATATCACAGCACTGTGTGGG	2593
Db	2533	AGTATCTTAGTTTGAATTACGATCAATTTAATCGAAACATTTCAAGTTTCGGGAGGGGGA	2592

QY	2594	GCTCTGTT-----TTTGTGATATATATGCCAACCAATCTCTGCAGAGGGCTGAGTTAA	2647
Db	2593	GTGTTGATTTACACACITCTCGCTCATCTCTACGTCCTCAAAACCCCGGTGTAGTTATAA	2652
QY	2648	AAATGAGTGAATTAATATCTCTAACGGGCTAATTTTACCTTAAATTCCTCAATGTTCCG	2707
Db	2653	ATTTCTAAATACITTAATGTTTCAACAGGTCAGTTTAAGATTTAAAACTTTCAGGCTCAA	2712
QY	2708	GGGATGACGCTTTTAAATCAACAAAGACTTAACCATAAATGCAACCAATTTCAAAATTCA	2767
Db	2713	CAAAAACTGGCTTCTCAATAGAGAAAGATTTAACTTTAAATGCCACC-----GGAGGCA	2766
QY	2768	GGCTCAGACAGAGGAAGATGATTTTATGACGGGTACGCACGCAATGCCATCAATTCAA	2827
Db	2767	ACATACACTTTTGGCAAGTTGAAGGCCGCGATGGAATGATTTGGTAAAGGCATTTGACCA	2826
QY	2828	CCTACACATATCCATCTCTGGGCGTAAATGTCACCCCTTGGTGGACAAAACCTCAAGCAGCA	2887
Db	2827	AAAAAAACATAACCTTTGAAGGAGGTAAACATCACCTTTGGCTCCAGAAAGCGTAAACAG	2886
QY	2888	GCATTACGGGGAATATTACTATTCGAGAAAGCAGCAAAATGTTTACGTAGAGCCAAATAACG	2947
Db	2887	AAATCGAAGGCAATGTTACTATCAATAACAAGCTAACGCTACTCTTATCGGTCGGAATT	2946
QY	2948	CCCCTAATCAGCAAAACATAAGGGATAGAGTTATATAAACTTTGGCAGCTTCTCGTTAAATG	3007
Db	2947	TTGACAACCATCAAAA-----ACCTTTAACTATTAAANAAGATGTCATTAATAGCG	3000
QY	3008	GGAGTTTAAAGTTAACTGGCGAAAATGCAAGATATTAAAGGCAAATCTCACTATTTCAGAAA	3067
Db	3001	GCACCTTTACCGCTCGAGGCAATATGTCAATATAGCCGGAATCTTACCGTTGAAAGTA	3060
QY	3068	GGCCCACTTTTAAAGGAAAGACTAGAGATACCCTAAATATACCCGGCAATTTTACCAATTA	3127
Db	3061	ACGCTAATTTCAAAGCTATCACAAAATTCACITTTTAATCTAGCGGCTTGTTTGACACACA	3120
QY	3128	ATGGCACTGCCGAATTAATTAACACAAAGAGTGGTAAAACTTTGGCAATGTTACCAATG	3187
Db	3121	AAGGCAATTCAAATATTTCCATTTGCCAAGGAGGGGCTCGCTTTTAAAGACATTTGATAATT	3180
QY	3188	ATGTGTGATTTAAACATTTACCCTACGCTTAACGCAACCAAGAGCATCATCGCGGAG	3247
Db	3181	CCAAGATTTAAGCATCACCACTCCAGCTCCACTTTACCGCACTATTATAAGCGGCA	3240
QY	3248	ATATTAATCAACAAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCTGAAATCC	3307
Db	3241	ATATAACCAATAAAAACGGTATTTTAAATATTACGAACGAAGTAGTACTGAAATGC	3300
QY	3308	AAATTTGGCGGCAATATCTCGCAAAAAGAGCAACCTCAGGATTTCTTCCGATATAAATTA	3367
Db	3301	AAATTTGGCGGGATCTCTCGCAAAAAGAGGTAATCTCAGGATTTCTTGTGACAAAAATCA	3360
QY	3368	ATATCACCAACAGATAACAATCAAAAAGGTTATGATGGAGAGGACTCTAGTTTCAGATG	3427
Db	3361	ATATTACCAACAGATAACAATCAAGGCAGGTGTTGATGGGAGAAATTCCTCGATTTCAGAGC	3420
QY	3428	CGACAAGTAATGCCCACTTAATTAACCAAGAATTCGAAATGACAGAAGCACTAA	3487
Db	3421	CGACAACAATGCCCAATCTAACCATTTAAACCAAGAATTTGAAATTAACGGAAGCACTAA	3480
QY	3488	GTATTTCCAGGTTTCAATAAAGCAGAGATTACAGCCAAAAGATGGTAGAGATTTTAACATTG	3547
Db	3481	ATATTTCCAGGTTTCAATAAAGCAGAGATTACAGCTAAAGATGGTAGTCAATTTAACTATTG	3540
QY	3548	GCACAGTAATGACGGTTAACAGCGGTGCCGAAGCCAAAACACTTAATTTTAACATGTTA	3607
Db	3541	GTAACACCAATAGTGTCTGA---TGGTACTAATGCCAAAAAAGTAACCTTTTAAACCAAGTTA	3597
QY	3608	AAGATTCAAAAATCTCTGCTGACGCTCACATGTGACACTAAATAGCAAAAGTGAACACAT	3667
Db	3598	AAGATTCAAAAATCTCTGCTGACGCTCACAAGGTGACACTACACAGAAAGTGAACACAT	3657
QY	3668	CTAGCAGCAATGGCGGACGTGAAAAGCAATAGCGACAACGATACCGGCTTAATCTATTACTG	3727


```

Db 3658 CCGGTAGTAATAACCAACTGAACATAGCAGTGACAAATAATGCGCGCTTAACATATCGATG 3717
QY 3728 CAAAAATGTAGAAGTAACAAAGATATTTACTTCTCAAAACAGATAAATATCACCAGGT 3787
Db 3718 CAAAAATGTAGAAGTAACAAAGATATTTACTTCTCAAAACAGATAAATATCACCAGGT 3777
QY 3788 C---GGAAAAGGTTACCAACAGAGGCTCGACCATTAACGCAACAAATGSCAA----- 3839
Db 3778 CAAAGTGAGAAATTTACCACTAAACAGAGTACAAACATTAACGCAACCACTGGTAACAGTGG 3837
QY 3840 ----- 3839
Db 3838 AGATAACCGCTCAAAACAGGTAGTATCTAGGTGGAATTCAGTCCAGCTCTGGCTCTGTAA 3897
QY 3840 ----- 3839
Db 3898 CACTTACTGCAACCGAGGCGCTCTTGCTGTAAAGCAATATTCGGGCAACACCGTTACTG 3957
QY 3840 ----- 3839
Db 3958 TTACTGCAATAGCGGTGCAATTAACACATTTGGCAGGCTCTACAATTAAGGAACCGAGA 4017
QY 3840 -AGCAAGTATTACAAACAAACAGGTGATATCAGCGGTACGATTTCCGGTAAACACGGTAA 3898
Db 4018 GTGTAACCACTTCAAGTCAATACAGCGATCGCGGTACGATTTCTGGTGACACAGTAG 4077
QY 3899 GTGTTAGCGGCTGCTGATTTAAACCACTAATTCGGGCTCAAAAATTTGAAGCGAAATCGG 3958
Db 4078 AGGTTAAAGCAACCGAAAGTTTAAACCACTCAATTCCAATTAAGGAATTAAGGAACCAACAG 4137
QY 3959 GTGAGGCTAATGTAAAGTGTCAACAGGTACAAATTCGGGCTCAAAAATTTCCGGTAAATACGG 4018
Db 4138 GCGAGGCTAAGTAAACAGTGTCAACAGGTACAAATTCGGGCTCAAAAATTTCCGGTAAATACGG 4197
QY 4019 TAAATGTTACGCAACCGCTGCGGATTTAAACAGTTGGGATTCGGGATTCGGGATTAATGCGA 4078
Db 4198 TAAATGTTACGCAACCGCTGCGGATTTAAACAGTTGGGATTCGGGATTCGGGATTAATGCGA 4257
QY 4079 CAGAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGAGTACTGACGCGGTTCTTA 4138
Db 4258 CAGAGGAGCTGCAACCTTAACCTACATTCGCGCAATTAACCTACGAGGCTAGTTTAC 4317
QY 4139 GCATCACTTCACTAAGGCTGAGTACAGCTCTTGCTCAGATGCTAGCATGCGCAGGAA 4198
Db 4318 ACATTAATGCAACCGGCTGAGTAAATCTTTCAGCTCAGGATGATGCTGCGAGGAA 4377
QY 4199 GCATTAATGCTGTAATGTGACATTAATATCAGGACCTTAAACACCGTGGCAGGCT 4258
Db 4378 GTATTAATGCGGCAATGTGACATTAATATCAGGACCTTAACTACCGTGAAGGGTT 4437
QY 4259 CGGATATTAAGCAACCGGCACTTGGTTATTAACGCAAAAGATGCTAAGCTAAATG 4318
Db 4438 CAAACATTAATGCAACCGGCTACCTTGGTTATTAACGCAAAAGAGCTGAGCTAAATG 4497
QY 4319 GTGATGATCAGGTGATGTACAGAGTGAATGCAAGTCAACGCAAGCGGCTCTGGTAGTG 4378
Db 4498 GCGCAGCATTTGGTGAACACACAGTGTGTAATGCAACCAACGCAAAATGGCTCGCGGACGC 4557
QY 4379 TGACTGCGGCAACCTCAAGCAGTGTGAATATCAGTGGGATTTAAACACAGTAAATGGGT 4438
Db 4558 TAATCGGCAACCTCAAGCAGTGAATATCAGTGGGATTTAAATCACAATAATGGAT 4617
QY 4439 TAAATATCATTTCAAGATGTTGAACACATCTGCTGCTTAAAGGCAAGCAATTCGAG 4498
Db 4618 TAAATATCATTTCAAGATGTTGAACACATCTGCTGCTTAAAGGCAAGCAATTCGAG 4677
QY 4499 TGAATATATCCAGCGGTGTAGCAAGTGTAGAAGAAAGTAATTTGAAGCAAGCGCTCC 4558
Db 4678 TGAATATATTTCAACCGGTTAGCAAGCGTGTAGTAAATTTGAAGCAAGCGCTCC 4737
QY 4559 TTGAAAAAGTAAAGATTTATCTGATGAAGAAAGCAACATTAAGCTTAACTGGTGTAA 4618

```

```

Db 4738 TTGAGAAGGTAAAGAGATTTATCTGCTGAAGAAAGAGAGCGTTAGCTAAACTTTGGAGTAA 4797
QY 4619 GTGCTGTACGTTTGTGTTGAGCCAAATATACAAATACAGTCAATACACAAAATGAATTTA 4678
Db 4798 GTGCTGTACGTTTGTGTTGAGCCAAATATACAAATACAGTCAATACACAAAATGAATTTG 4857
QY 4679 CAACGACACCGTCAAGTCAAGTGAATTTCTGAAGGTGAAGGCGTCTCTCAAGTGGTA 4738
Db 4858 CAACGACACCGTCAAGTCAAGTGAATTTCTGAAGGTGAAGGCGTCTCTCAAGTGGTA 4917
QY 4739 ATGCGGCACGAGTATGATACCAATTTCTGACGATGAGGACCGCTGATCAGTAATTCACA 4798
Db 4918 ATGCGGCACGAGTATGATACCAATTTCTGATATCTGATACGCGGTGATGATTAATTCACA 4977
QY 4799 AGGTAGATTTTCATCTCTGCAATGAAGTCAATTTATTTTCGTAATTTTACTGTGTGGGTTA 4858
Db 4978 AGGTAGATTTTCATCTCTGCAATGAAGTCAATTTATTTTCGTAATTTTACTGTGTGGGTTA 5037
QY 4859 AAGTTTCAGTACGCGGCTTTTACCCATCTTGTAAAAAATACGAGAGTAAATCAATAAAAATTT 4918
Db 5038 AAGTTTCAGTACGCGGCTTTTACCCATCTTGTAAAAAATACGAGAGTAAATCAATAAAAATTT 5097
QY 4919 TTAACAGGTTTATTTATG 4937
Db 5098 TTAACAGGTTTATTTATG 5116

RESULT 11
AAAS2195
ID AAAS2195 standard; DNA; 5116 BP.
XX
AC AAAS2195;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain 12 hmw1A gene, SEQ ID NO:66.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; Nthi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis; ds.
XX
OS Haemophilus influenzae strain 12.
XX
Key Location/Qualifiers
CDS 351..4961
FT /*tag= a
FT /product= "Haemophilus influenzae strain 12 hmw1A
FT protein"
XX
WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
WPI: 2000-303789/26.
XX
P-PSDB; AAB01846.
XX
Nucleic acid molecule for producing recombinant high molecular weight
proteins of Haemophilus which are used as a vaccine to provide
protection against Haemophilus induced diseases in humans.
XX
Example 16; Fig 28A-Q; 307pp; English.
XX

```

CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in *Escherichia coli*. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in *E. coli* (e.g., the T7 promoter) operably linked
 CC to a modified hmWABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene
 CC clusters termed hmw1ABC and hmw2ABC. Each hmWABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmWABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMW. The invention also discloses hmwA genes (AA52175-A52198)
 CC and HMW proteins (AA501824-B01849) from the non-typeable H. influenzae
 CC strains Jovc, K1, K21, LDC2, FWH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an hmwA gene from a non-typeable strain of H.
 CC influenzae.

XX
 SQ Sequence 5116 BP; 1766 A; 1008 C; 1037 G; 1305 T; 0 other;

Query Match 54.3%; Score 2680.2; DB 21; Length 5116;
 Best Local Similarity 73.5%; Pred. No. 0;
 Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps 16;

QY 41 ATGCAAAACACAAATACAAACCTTTTTCAGCTCTATATGCAATATTTTAAAAAAT 100
 DB 41 ATGCAAAACACAAATACAAACCTTTTTCAGCTCTATATGCAATATTTTAAAAAAT 99
 QY 101 AGTAAATTCGCCCATATAAATGGTAAATCTTTCATCTTTCATCTTTCATCTTCATC 160
 DB 100 AGTAAATTCGCCCATATAAATGGTAAATCTTTCATCTTTCATCTTTCATCTTCATC 159
 QY 161 TTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCT 220
 DB 160 TTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCT 219
 QY 221 ATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 280
 DB 220 ATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 279
 QY 281 GAATGAAGAGGAGCTGAACGACGCAATGATAAGTAAATTTTAACTAACT 340
 DB 280 GAATGAAGAGGAGCTGAACGACGCAATGATAAGTAAATTTTAACTAACT 339
 QY 341 TAGGAGAAATATGACAAAGATATATCTCTCAAAATTCAGCAACGCTCAATGCTTTGG 400
 DB 340 TAGGAGAAATATGACAAAGCTATATCTCTCAAAATTCAGCAACGCTCAATGCTTTGG 399
 QY 401 TTGCTGTGCTGAATTTGGCAGCGGGTTGTGACCAATTTCCACAGAAAAAGGCTTCGCGTAT 460
 DB 400 TTGCTGTGCTGAATTTGGCAGCGGGTTGTGACCAATTTCCACAGAAAAAGGCGAAGAC 459
 QY 461 TTATATCTTTAGTGTACCACTTAGCGTTAAAGCCACTTTCGCTATGTTACTATCTT 520
 DB 460 TTGCTGTGCTGAATTTGGCAGCGGGTTGTGACCAATTTCCACAGAAAAAGGCGAAGAC 519
 QY 521 TAGGTGTAACATCTATTCACAATCTGTTTTAGCAAGCGGCTTACAGGAAGTGGATGTAG 580
 DB 520 TAGGTGTAACATCTATTCACAATCTGTTTTAGCAAGCGGCTTACAGGAAGTGGATGTAG 579
 QY 581 TACAGGACAGCCACTATGCAAGTAGATGGTAATAAACCATTTATCCGCAACAGTGTG 640
 DB 580 TACAGGACAGCCACTATGCAAGTAGATGGTAATAAACCATTTATCCGCAACAGTGTG 639

QY 641 ACGTATCATTAATTTGGAACAATTTTACATCGACCAAAATGAATGGTGCAAGTTTTTAC 700
 DB 640 ACGTATCATTAATTTGGAACAATTTTACATCGACCAAAATGAATGGTGCAAGTTTTTAC 699
 QY 701 AAGAAACAAACAACTCCGCGGTATTCACACCGTGTACATCTAACCAAAATCTCCCAATTA 760
 DB 700 AAGAAACAAACAACTCCGCGGTATTCACACCGTGTACATCTAACCAAAATCTCCCAATTA 759
 QY 761 AAGGGATTTAGATCTTAACGGCAAGCTTTTTTAATCAACCCAAATGATATCAATAG 820
 DB 760 AAGGGATTTAGATCTTAACGGCAAGCTTTTTTAATCAACCCAAATGATATCAATAG 819
 QY 821 GTAAGACGCAATTTAATCAACATTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 880
 DB 820 GTAAGACGCAATTTAATCAACATTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 879
 QY 881 AATACATCAAGCGCGGTAAATTTTACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAA 940
 DB 880 AATACATCAAGCGCGGTAAATTTTACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAA 939
 QY 941 TTGTGAATCAAGCGGTAAATTTTACTGTGTAAGACGCAAGCTGTAAATCTTATTTGGTGGCA 1000
 DB 940 TTGTGAATCAAGCGGTAAATTTTACTGTGTAAGACGCAAGCTGTAAATCTTATTTGGTGGCA 999
 QY 1001 AAGTGAACAAACGAGGGTGTGATTTAGCTGTAATGTGCGACGATTTCTTTACTCGCAGGCG 1060
 DB 1000 AAGTGAACAAACGAGGGTGTGATTTAGCTGTAATGTGCGACGATTTCTTTACTCGCAGGCG 1059
 QY 1061 AAAAATCACCATCAGCGATATAATTAACCCCAACCATTTACTTACAGCAATTCGCGCGCTG 1120
 DB 1060 AAAAATCACCATCAGCGATATAATTAACCCCAACCATTTACTTACAGCAATTCGCGCGCTG 1119
 QY 1121 AAAATGAAGCGGTCAATCTGGCGGATATTTTGCACAAAGCGGTAAACATTAATGTCCGCTG 1180
 DB 1120 AAAATGAAGCGGTCAATCTGGCGGATATTTTGCACAAAGCGGTAAACATTAATGTCCGCTG 1179
 QY 1181 CTGCCACTTTCGAAACCAAGGTAACTTCTGCTGATTTCTGTAAGCAAAAGATAAAGCG 1240
 DB 1180 CTGCCACTTTCGAAACCAAGGTAACTTCTGCTGATTTCTGTAAGCAAAAGATAAAGCG 1239
 QY 1241 GCAATATTTGTTCTTTCGCGCAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTC 1300
 DB 1240 GCAATATTTGTTCTTTCGCGCAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTC 1299
 QY 1301 AAAATCAGCAAGCTTAAGCGCGCAAGCTGATTTACAGCGGATAAAAGTACACATTAATAA 1360
 DB 1300 AAAATCAGCAAGCTTAAGCGCGCAAGCTGATTTACAGCGGATAAAAGTACACATTAATAA 1359
 QY 1361 CAGGTGCAAGTTATCGACCTTTTCAGTAAAGAGGGGCAAACTTACCTTTGGCGGTGACG 1420
 DB 1360 CAGGTGCAAGTTATCGACCTTTTCAGTAAAGAGGGGCAAACTTACCTTTGGCGGTGACG 1419
 QY 1421 AGCGCGGCAAGTAAACCGCATTTCAATTTAGCAAGAAACCTCTTTTAAAGAAAGCT 1480
 DB 1420 AGCGCGGCAAGTAAACCGCATTTCAATTTAGCAAGAAACCTCTTTTAAAGAAAGCT 1479
 QY 1481 CAACCATCAATGTATCAGCGCAAAACAAAGAGCGGCGCTTATTTGTTGGGGGATATTG 1540
 DB 1480 CAACCATCAATGTATCAGCGCAAAACAAAGAGCGGCGCTTATTTGTTGGGGGATATTG 1539
 QY 1541 CGTTAATTGACGCAATATTAAACGCTCAAGGTAGTGGTGATATCGCTTAAACCGGTGTT 1600
 DB 1540 CGTTAATTGACGCAATATTAAACGCTCAAGGTAGTGGTGATATCGCTTAAACCGGTGTT 1599
 QY 1601 TTGTGGAGACGCGGGGCAATTTATTTATCCATTTGACAGCAATGCAATTTTAAACAAAG 1660
 DB 1600 TTGTGGAGACGCGGGGCAATTTATTTATCCATCAAGACAAATGCAATTTTGGAGCGCAAG 1659
 QY 1661 AGTGGTTGCTAGACCTCTGATGATCAACAAATTAAGCCGAGACCCCTTCCCAATAATA 1720
 DB 1660 AGTGGTTGTTAGACCCGATATGATCTATTTAATGACAGAAACAGCAGCAGCAATA 1719
 QY 1721 CCGGTATAAATGATGAATTTCCCAACAGCGCGGTGAAGCAAGCACCCTTAAAAAATA 1780

[illegible]

QY 3840 ----- 3839
Db 3898 CACTTACTGCAACCGAGGCGCTCTGCTGTAAGCAATATTTTCGGGCAACACCGTTACTG 3957
QY 3840 ----- 3839
Db 3958 TTACTGCAATATACCGGTGCATTAACCACTTTGGCAGGCTCTACAAATTAAGGAACCGAGA 4017
QY 3840 -AGCAAGTATTACAAACCAACAGGTGATATCAGGGGTACGATTTCCGGTAAACCGGTAA 3898
Db 4018 GTGTAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTGGTGCACAGTAG 4077
QY 3899 GTGTTACGGGCACTGGTATTTTAACCACTAAATCCGGCTCAAAAATTAAGGGAATCGG 3958
Db 4078 AGGTTAAAGCAACCGAAGTTTAACCACTCAATTCGAATTTCAAAAATTAAGGCAACACAG 4137
QY 3959 GTGAGGCTAATGTAAAGTCAACAGGTACAATTTGGCGGTACAAATTTCCGGTAAATACGG 4018
Db 4138 GCGAGGCTAAGTCAACAGTCAACAGGTACAATTTGGGTGATGATTTCCGGTAAATACGG 4197
QY 4019 TAAATGTTACGGCAACCGTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATCGGA 4078
Db 4198 TAAATGTTACGGCAACCGTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATCGGA 4257
QY 4079 CAGAAGGAGTCAACCTTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTA 4138
Db 4258 CAGAAGGAGTCAACCTTTAACCTATCATCGGCAATTAACCTCCGAAGTGTACAC 4317
QY 4139 GCATCACTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATTCGACAGAA 4198
Db 4318 ACATTACTTCAACCAAGGGTCAGGTAAATCTTTCAGCTCAGGATGGTAGCGTTGACAGAA 4377
QY 4199 GCATTATGCTCTAATGTGACATTAATTAACACTACAGGCCTTTAACCCCGTGGCAGGCT 4258
Db 4378 GTATTATGCCCCAAATGTGACATTAATTAACACTACAGGCCTTTAACCCCGTGGCAGGCT 4437
QY 4259 CGGATTAATGAACAACAGCGGCACCTTGGTTTATTAACGCAAAAGATGCTAAGTAATG 4318
Db 4438 CAACATTAATGAACAACAGCGGTACCTTGGTTTATTAACGCAAAAGATGCTAAGTAATG 4497
QY 4319 GTGATCATCATGCTAGTAGTACAGAAATGATGCAATGCACTGCAACGCGCTGCTGGTAGT 4378
Db 4498 GCGCAGCATTTGGTAAACACAGATGCTGTAATGCAACACGCAATGGCTCCGGCAGCG 4557
QY 4379 TGACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGT 4438
Db 4558 TAATCGCGCAACCTCAAGCAGATGAACATCACTGGGGATTTAAATCAATAAATGGAT 4617
QY 4439 TAAATATCAATTCGAAAGATGGTAGAAACACTGTGCGCTTAAGAGCAAGGAATTCAGG 4498
Db 4618 TAAATATCAATTCGAAAGATGGTAGAAACACTGTGCGCTTAAGAGCAAGGAATTCAGG 4677
QY 4499 TGAATATATCCAGCCAGGTGTAGCAAGGTAGAAAGTAAATTTGAAGCGAAAGCGGTCC 4558
Db 4678 TGAATATATCAATTCGAAAGATGGTAGCAAGGTAGAAAGTAAATTTGAAGCGAAAGCGGTCC 4737
QY 4559 TTCAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAACTTGGGTGTA 4618
Db 4738 TTGAGAGTAAAGATTTATCTGATGAAGAAAGAGAAAGCGGTAGCTAACTTGGAGTAA 4797
QY 4619 GTCTGTACGTTTTGTTAGCCAAATTAACAATTTACAGTCAATACACAAATGAATTTA 4678
Db 4798 GTCTGTACGTTTTTATGAGCCAAATTAACAATTTACAGTCAATACACAAATGAATTTA 4857
QY 4679 CAACAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTA 4738
Db 4858 CAACAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTA 4917
QY 4739 ATGGCGCAGAGTGTACCAATTTGCTGACGATGACAGCGGTAGTCAGTAAATTTGACA 4798
Db 4918 ATGGCGCAGAGTGTGCGTTATATCGCTGATACCGCGGTAGTCAGTAAATTTGACA 4977

QY 4799 AGGTAGATTTTCATCTGCAATGAAGTCAATTTATTTTCGTATTTACTGTGTGGTTA 4858
Db 4978 AGGTAGATTTTCATCTGCAATGAAGTCAATTTATTTTCGTATTTACTGTGTGGTTA 5037
QY 4859 AAGTTCAGTACGGGCTTTACCCATCTTGTAATAAAATTTACGAGAAATACAAATTAATTT 4918
Db 5038 AAGTTCAGTACGGGCTTTACCCATCTTGTAATAAAATTTACGAGAAATACAAATTAATTT 5097
QY 4919 TTAACAGGTTATTTATG 4937
Db 5098 TTAACAGGTTATTTATG 5116
RESULT 12
AAQ49508
ID AAQ49508 standard; DNA: 9220 BP.
XX AC AAQ49508;
XX DT 26-APR-1994 (first entry)
XX DE Gene cluster for high molecular weight protein 1 (HMM1).
XX KW HMM: high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae; gene cluster; ss.
XX OS Haemophilus Influenzae.
XX FH Key
FT CDS Location/Qualifiers
FT 351..4958
FT /*tag= a
FT /product= High molecular weight protein 1.
FT 5114..6748
FT /*tag= b
FT /note= "One of a gene cluster for high molecular
FT weight protein 1"
FT CDS 7062..9011
FT /*tag= c
FT /note= "One of a gene cluster for high molecular
FT weight protein 1"
FT repeat_unit 2351..2380
FT /*tag= d
FT /label=
FT /note= "Possible error in sequence. Alternative
FT sequence for this region is CTTAAATGTT
FT TCCGAGAGTG CCGAGTTTAA (See AAR41725)"
XX PN W09319090-A.
XX PD 30-SEP-1993.
XX PF 16-MAR-1993; 93WO-US02166.
XX PR 16-MAR-1992; 92GB-0005704.
XX PA (BARE/) BARENKAMP S J.
XX PI (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX Barenkamp SJ;
DR WPI: 1993-320683/40.
DR P-PSDB; AAR41725, AAR41726, AAR41727.
XX High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
XX Claim 9; Figure 6; 100pp; English.
XX The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.

XX	Sequence 9220 BP; 3101 A; 1861 C; 1805 G; 2453 T; 0 other;	
SQ	Query Match 54.2%; Score 2677; DB 14; Length 9220; Best Local Similarity 73.4%; Pred. No. 0; Matches 3759; Conservative 0; Mismatches 1095; Indels 265; Gaps 16;	
QY	41 ATGACAAACAACAAATACAAACACCTTTTTCGAGCTATATGCAAAATATTTTAAAAAAT 100	Db 1000 AAGTGAACAAACGAGGGTGTGATTAGCGTAATAGTGGCAGCAATTTCTTTACTCGCAGGC 1059
DB	41 ATGACAAACAACAAATACAAACACCTTTTTCGAGCTATATGCAAAATATTTTAAAAAAT 99	QY 1061 AAAAAATCACCATCAGCGATATATTAACCAACCAATTAATTAACGATTTGCGCGCGCTG 1120
QY	101 AGTATAATCCGCCATATAAATGATTAATCTTTTCATCTTTTCATCTTTTAAATCTTTCATC 160	Db 1060 AAAAAATCACCATCAGCGATATATTAACCAACCAATTAATTAACGATTTGCGCGCGCTG 1119
DB	100 AGTATAATCCGCCATATAAATGATTAATCTTTTCATCTTTTCATCTTTTAAATCTTTCATC 159	QY 1121 AAAAAATCACCATCAGCGATATATTAACCAACCAATTAATTAACGATTTGCGCGCGCTG 1180
QY	161 TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220	Db 1120 AAAAAATCACCATCAGCGATATATTAACCAACCAATTAATTAACGATTTGCGCGCGCTG 1179
DB	160 TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219	QY 1181 CTGCCACTATTCGAAACCAAGGTAAGCTTTCTGCTGATTTCTGTAAGCAAGATATAAAGCG 1240
QY	221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 280	Db 1180 CTGCCACTATTCGAAACCAAGGTAAGCTTTCTGCTGATTTCTGTAAGCAAGATATAAAGCG 1239
DB	220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 279	QY 1241 GCAATATTTGTTCTTCGCGCAAGAGGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTC 1300
QY	281 GAATGAAGAGGAGCTGAACGAACCAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGG 340	Db 1240 GCAATATTTGTTCTTCGCGCAAGAGGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTC 1299
DB	280 GAATGAAGAGGAGCTGAACGAACCAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGG 339	QY 1301 AAAAAATCACCATCAGCGATATATTAACCAACCAATTAATTAACGATTTGCGCGCGCTG 1360
QY	341 TAGGAGAAATATGAACAGATATATGCTCTCAATTCAGCAACGCGTGAATCTTTGG 400	Db 1300 AAAAAATCACCATCAGCGATATATTAACCAACCAATTAATTAACGATTTGCGCGCGCTG 1359
DB	340 TAGGAGAAATATGAACAGATATATGCTCTCAATTCAGCAACGCGTGAATCTTTGG 399	QY 1361 CAGGTGCAGTATTCGACCTTTCAGGTAAAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGG 1420
QY	401 TTGCTGTCTGAATTTGCGAGGGTGTGACCAATTCACAGAAAGAGGCTTCGCGCTATG 460	Db 1360 CAGGTGCAGTATTCGACCTTTCAGGTAAAGAGGGGAGGAGGAGGAGGAGGAGGAGG 1419
DB	400 TTGCTGTCTGAATTTGCGAGGGTGTGACCAATTCACAGAAAGAGGAGGAGGAGGAGGAGG 459	QY 1421 AGCGCGGAGAGGTAAAGCGGATTCATTTAGCAAGAAACCTCTTTAGAAAAAGGCT 1480
QY	461 TTACTATCTTTAGGTGTAAACCACTTAGCGTTAAAGCCACTTTTCGCGTATGTACTACTT 520	Db 1420 AGCGCGGAGAGGTAAAGCGGATTCATTTAGCAAGAAACCTCTTTAGAAAAAGGCT 1479
DB	460 TTACTATCTTTAGGTGTAAACCACTTAGCGTTAAAGCCACTTTTCGCGTATGTACTACTT 519	QY 1481 CAACCATCAATGTATCAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 1540
QY	521 TAGGTGTAAACATCTTTTACCAATCTTTTAGCAAGCGGCTTACAGGAATGGATGTAG 580	Db 1480 CAACCATCAATGTATCAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 1539
DB	520 TAGGTGTAAACATCTTTTACCAATCTTTTAGCAAGCGGCTTACAGGAATGGATGTAG 579	QY 1541 CGTTAATTTAGCGCAATATTAACGCTCAAGGTAGTGTGATATCGCTAAAAACCGGTGTT 1600
QY	581 TACAGCGCACAGCCACTATGCAAGTAGATGGTAATAAACCATTATCGCAACAGTGTG 640	Db 1540 CGTTAATTTAGCGCAATATTAACGCTCAAGGTAGTGTGATATCGCTAAAAACCGGTGTT 1599
DB	580 TACAGCGCACAGCCACTATGCAAGTAGATGGTAATAAACCATTATCGCAACAGTGTG 639	QY 1601 TTGTGAGACATCGGGGCAATTTATTCATTCAGCAAGCAATGCAATTTGTGACGCCAAG 1660
QY	641 ACCTATCATTAATTTGGAACAATTTACATCGACCAAAATGAATGTCAGTGTGTTTAC 700	Db 1600 TTGTGAGACATCGGGGCAATTTATTCATTCAGCAAGCAATGCAATTTGTGACGCCAAG 1659
DB	640 ACCTATCATTAATTTGGAACAATTTACATCGACCAAAATGAATGTCAGTGTGTTTAC 699	QY 1661 AGTGTGTGTAGACCTGTATGATTAACCAATTCAGCGGAGGAGGAGGAGGAGGAGGAGG 1720
QY	701 AAAAAACAACCTCCGCGTATTCACCGTGTACATCTAACCAATTCGCAATTCGCAATTA 760	Db 1660 AGTGTGTGTAGACCTGTATGATTAACCAATTCAGCGGAGGAGGAGGAGGAGGAGG 1719
DB	700 AAAAAACAACCTCCGCGTATTCACCGTGTACATCTAACCAATTCGCAATTCGCAATTA 759	QY 1721 CCGGTATAATGATTAATTCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1780
QY	761 AAGGATTTTATCTAACGCAAGCTTTTAAATCAACCCAAATGGTATCAACAATAG 820	Db 1720 CCGGTATAATGATTAATTCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1779
DB	760 AAGGATTTTATCTAACGCAAGCTTTTAAATCAACCCAAATGGTATCAACAATAG 819	QY 1781 GCGAATCAAAACAGCGTAAACCAATTAACCAATTTTCAATTTTCAATTTTCAATTTTCA 1840
QY	821 GTAAGAGCGCAATTTAATCACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 880	Db 1780 AAGAA---AAGAAACAATTAACCAACCAATTTTCAATTTTCAATTTTCAATTTTCA 1836
DB	820 GTAAGAGCGCAATTTAATCACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 879	QY 1841 CAATGAATTAACCGCATCAAGAAACCTTACCGTAAATAGCTCAATCAACATCGGAAGA 1900
QY	881 AAAACATCAAGCGCGTAAATTTACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAA 940	Db 1837 TTGTTAATCACTCACTTAATCAAGCAATCTATGTCATAGCTCAATTTTCAATTTTCA 1893
DB	880 AAAACATCAAGCGCGTAAATTTACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAA 939	QY 1901 ACTCCCACTTAATTTCCATAGTAAAGGTCAGGTCAGGAGGAGGAGGAGGAGGAGGAGG 1960
QY	941 TTGTGAATCAGGTTTAAATCTGCTGCTGAAGAGCGGATGTAATCTTATTGGTGGA 1000	Db 1894 ATGCAAGCTTAATCTTTGAGTGGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1953
DB	940 TTGTGAATCAGGTTTAAATCTGCTGCTGAAGAGCGGATGTAATCTTATTGGTGGA 999	QY 1961 ATATTAAC-----TTCTAAAGCGGAAATTAACCAATTTTCTGCGCGATGGG 2008
QY	1001 AAGTGAACAAACGAGGGTGTGATAGGCTAAATGCTGAGCAATTTCTTTTACTCGCAGGC 1060	Db 1954 ATATTAACCGCGTGTATGATACCGAGGAGTGAACCAATTTTCAATTTTCTGCGCGCTGGG 2013


```

Db 4258 CAGAGGAGCTCAACCTTAACATACATATCGGCAAAATTAACACCGAGTAGTTAC 4317
Qy 4139 GCATCAGTCAACTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGATAGTCAGGAA 4198
Db 4318 ACATTACTTCAGCAGGTCAGGTAAATCTTTCAGCTCAGATGATGTCAGGTGAGGAA 4377
Qy 4199 GCATTAATGCTGCTAATGTCAGATTAATACATACAGGCACTTAACACCGTGGCAGCT 4258
Db 4378 GTATTAAATGCCCAATGTCAGCTAAATACATACAGGCACTTAACACCGTGGGTT 4437
Qy 4259 CGGATATTAAGCAACACCGGCTGCTGGTATTAAAGCAAAAGATGCTAAGCTAAATG 4318
Db 4438 CAACATTAATGCAACACCGGCTGCTGGTATTAAAGCAAAAGATGCTAAGCTAAATG 4497
Qy 4319 GTGATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAACGCAAGCGGCTGCTGATG 4378
Db 4498 GCGCAGCATTTGGTAAACACAGTGGTAAATGCAACCAACGCAAAATGGCTCCGCGAGCG 4557
Qy 4379 TGACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGT 4438
Db 4558 TAATCGCAGCAACCTCAAGCAGTGAATATCACTGGGATTTAAACACAGTAAATGGT 4617
Qy 4439 TAAATATCATTTGCAAGATGTTAGAACACATGTCGGTAAAGCAAGGAAATGGG 4498
Db 4618 TAAATATCATTTCAAAAACGTTATTAACACCGTACTGTTAAAGCGGTTAAATGGATG 4677
Qy 4499 TGAATATATTCAGGCGGCTGAGCAAGTGTAGAGAGTAAATGGAAGCAAGGAAATGGG 4558
Db 4678 TGAATATATCAACCGGCTGAGCAAGTGTAGAGAGTAAATGGAAGCAAGGAAATGGG 4737
Qy 4559 TTGAAGAAGTAAAGATTTATCTGATGAAGAAGAGAAACATTTAGCTAAACCTGGTGAA 4618
Db 4738 TTGAAGAAGTAAAGATTTATCTGATGAAGAAGAGAAACATTTAGCTAAACCTGGTGAA 4797
Qy 4619 GTGCTGTACGTTTGTGAGGCAAAATACAAATTTACAGTCAATACACAAAATGAATTTA 4678
Db 4798 GTGCTGTACGTTTGTGAGGCAAAATACAAATTTACAGTCAATACACAAAATGAATTTG 4857
Qy 4679 CAACAGACCGTCAAGTGAATTTCTGGAAGTAAAGCGGTTTCTCAAGTGTA 4738
Db 4858 CAACAGACCGTCAAGTGAATTTCTGGAAGTAAAGCGGTTTCTCAAGTGTA 4917
Qy 4739 ATGGCGCAGAGTATGATACCAATGTTGTCAGGATGAGCAGCGGAGTAAATGACA 4798
Db 4918 ATGGCGCAGAGTATGATACCAATGTTGTCAGGATGAGCAGCGGAGTAAATGACA 4977
Qy 4799 AGGTAGATTTTCATCTGCAATGAAGTCAATTTATTTCTGATTTTACTGTGGGTTA 4858
Db 4978 AGGTAGATTTTCATCTGCAATGAAGTCAATTTTCTGATTTTACTGTGGGTTA 5037
Qy 4859 AAGTTCAGTACGGCTTTACCCATCTTGTAAATAATACGAGAAATACAAATAAGATATT 4918
Db 5038 AAGTTCAGTACGGCTTTACCCATCTTGTAAATAATACGAGAAATACAAATAAGATATT 5097
Qy 4919 TTAACAGGTTATATTATG 4937
Db 5098 TTAACAGGTTATATTATG 5116

```

RESULT 13

ID AAT90996 standard; DNA; 9171 BP.

XX AAT90996;

XX 14-APR-1998 (first entry)

XX Non-typeable Haemophilus influenzae hmw1 gene cluster.

XX Non-typeable Haemophilus; high molecular weight surface protein;

KW Hmw1; hmw1 gene; immunogen; vaccine; otitis media; ss.

XX

```

OS Haemophilus influenzae strain 12.
XX
FH Key Location/Qualifiers
FT CDS 351..4958
FT /*tag= a
FT /*note= "hmw1 gene"
FT CDS 5114..6748
FT /*tag= b
FT /*note= "ORF-b"
FT CDS 7062..9011
FT /*tag= c
FT /*note= "ORF-c"
PN W09736914-A1.
XX
PD 09-OCT-1997.
XX
PF 01-APR-1997; 97WO-US04707.
XX
PR 01-APR-1996; 96US-0617697.
XX
PA (BARE/) BARENKAMP S J.
XX
PI Barenkamp S J;
XX
PI WPI; 1997-503038/46.
XX
DR P-PSDB; AAW30293.
XX
PT High molecular weight proteins of non-typeable Haemophilus
PT influenzae - useful for vaccine production
XX
PS Disclosure; Page 78-82; 183pp; English.
XX
CC This nucleic acid comprises a gene cluster for the hmw1 gene (see
CC also AAT90994) that encodes high molecular weight surface protein
CC Hmw1 (see AAW30293) of non-typeable Haemophilus influenzae strain 12.
CC In addition to the hmw1 gene there are 2 additional downstream
CC open reading frames that are required for the correct processing
CC and secretion of the hmw1 gene product. The ORF-b derived amino
CC acid sequence demonstrates similarity with the derived amino acid
CC sequences of 2 genes which encode proteins required for secretion
CC and activation of haemolysins of P. mirabilis and S. marcescens.
CC Hmw proteins (see AAW30291-94) can be used in vaccines, as immunogens
CC for preparation of antibodies and as antigens for detection of
CC these antibodies.
XX
SQ Sequence 9171 BP; 3085 A; 1855 C; 1794 G; 2437 T; 0 other;
Query Match 52.0%; Score 2565.4; DB 18; Length 9171;
Best Local Similarity 72.4%; Pred. No. 0;
Matches 3708; Conservative 0; Mismatches 1096; Indels 315; Gaps 17;
Qy 41 ATGACAAACAAATACAAACACCTTTTTCGAGTCTATATGCAAAATATTTTAAAAAAT 100
Db 41 ATGACAAACAAATACAAACACCTTTTTCGAGTCTATATGCAAAATATTTTAAAAAAT 99
Qy 101 AGTATAAATCCGCCATATAAATGCTATAATCTTTTCATCTTTTCACTTTTCATC 160
Db 100 AGTATAAATCCGCCATATAAATGCTATAATCTTTTCATCTTTTCACTTTTCATC 159
Qy 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220
Db 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219
Qy 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 280
Db 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 279
Qy 281 GAATGAAGAGGAGCTGAACGAAACCAATGATAAGTAATTTAATTTGTTCAACTAACCT 340
Db 280 GAATGAAGAGGAGCTGAACGAAACCAATGATAAGTAATTTAATTTGTTCAACTAACCT 339
Qy 341 TAGGAGAAATATGAACAAGATATATCGTCTCAAAATTCAGCAACCGCTGATGCTTTGG 400

```


|||||
Db 340 TAGGAGAAATATGACAAAGATATATCGTCTCAATTCAGCAACGCCCTGAATGTTTGG 399
QY 401 TTGCTGTCTGAATTTGGCAGCGGTTGTGACCATTTCCACAGAAAGGCTTCCCGCTATG 460
Db 400 TTGCTGTCTGAATTTGGCAGCGGTTGTGACCATTTCCACAGAAAGGCGCGGAAAC 459
QY 461 TTACTATCTTAGGTGAACCATTTAGCGTTAAAGCCATTTCCGCTATGTTACTATCTT 520
Db 460 CTGCTCGATGAAGTGGTCACTTAGCGTTAAAGCCATTTCCGCTATGTTACTATCTT 519
QY 521 TAGGTGAACATCTATCCACAATCTGTTTACAGCGGCTTACAGGAATGGATGTAG 580
Db 520 TAGGTGAACATCTATCCACAATCTGTTTACAGCGGCTTACAGGAATGGATGTAG 579
QY 581 TACAGCGCACAGCCCATGATGCAAGTAGATGTTAAATAAACCATTTACCGCAACAGGTTG 640
Db 580 TACAGCGCACAGCCCATGATGCAAGTAGATGTTAAATAAACCATTTACCGCAACAGGTTG 639
QY 641 AGCTATCATTAATTTGGAAACAATTTAAACATCGACCAAAATGAAATGGTGGAGTTTAC 700
Db 640 AGCTATCATTAATTTGGAAACAATTTAAACATCGACCAAAATGAAATGGTGGAGTTTAC 699
QY 701 AAGAAACAACTCCCGCTATTCAACCGTTTACATCTAACCAAACTCTCCCAATTA 760
Db 700 AAGAAACAACTCCCGCTATTCAACCGTTTACATCTAACCAAACTCTCCCAATTA 759
QY 761 AAGGATTTTATGATCTAACGGCAAGTCTTTTAAATCAACCCAAATGGTATCACAATAG 820
Db 760 AAGGATTTTATGATCTAACGGCAAGTCTTTTAAATCAACCCAAATGGTATCACAATAG 819
QY 821 GTAAGAGCAATTTAATCACTAATGGCTTTACGGCTTACAGCTAGACATTTCTAACG 880
Db 820 GTAAGAGCAATTTAATCACTAATGGCTTTACGGCTTACAGCTAGACATTTCTAACG 879
QY 881 AAAACATCAAGCGGCTAATTTACCTTTGAGCAACCAAGATAAAGCGCTCGCTGAA 940
Db 880 AAAACATCAAGCGGCTAATTTACCTTTGAGCAACCAAGATAAAGCGCTCGCTGAA 939
QY 941 TTGTGAATCAAGCTTTAATTTACTCTGGTAAAGCGGAGTGAATCTTATTTGGTGGCA 1000
Db 940 TTGTGAATCAAGCTTTAATTTACTCTGGTAAAGCGGAGTGAATCTTATTTGGTGGCA 999
QY 1001 AAGTGAAGGAGGCTGATTTAGCGTAAATGGTGGCAGATTTCTTTACTCGCAGGCG 1060
Db 1000 AAGTGAAGGAGGCTGATTTAGCGTAAATGGTGGCAGATTTCTTTACTCGCAGGCG 1059
QY 1061 AAAAATCACCATCAGCGATATATTAACCCACCATTTACTTTACAGCATTTGCGCGCTG 1120
Db 1060 AAAAATCACCATCAGCGATATATTAACCCACCATTTACTTTACAGCATTTGCGCGCTG 1119
QY 1121 AAAATGAAGCGGTCAATCTGGCGATATTTTGGCAAGGCGGTAAACATTAATGTCCGTG 1180
Db 1120 AAAATGAAGCGGTCAATCTGGCGATATTTTGGCAAGGCGGTAAACATTAATGTCCGTG 1179
QY 1181 CTGCGCATTTGAAACCAAGGTAAACTTTCTGCTGATCTGTAAAGCAAGATAAAGCG 1240
Db 1180 CTGCGCATTTGAAACCAAG----- 1200
QY 1241 GCAATATTTCTTTCCGCAAGAGGCTGAAGCGGAAATTTGGCGGTGAATTTCCGCTC 1300
Db 1201 -----CTTTCCGCAAGAGGCTGAAGCGGAAATTTGGCGGTGAATTTCCGCTC 1249
QY 1301 AAAATCAGCAAGTAAAGCGCGCAAGCTGATGATTACAGCGGATAAAGTCACATTAATA 1360
Db 1250 AAAATCAGCAAGTAAAGCGCGCAAGCTGATGATTACAGCGGATAAAGTCACATTAATA 1309
QY 1361 CAGGTGAGTTATCGACCTTTACAGTAAAGAGGGGGAACAACTTTACCTTGGCGGTGAG 1420
Db 1310 CAGGTGAGTTATCGACCTTTTACAGTAAAGAGGGGGAACAACTTTACCTTGGCGGTGAG 1369
QY 1421 AGCGGCGGAGGTAAAGCGCATTTCAATTTAGCAAGAAACCTCTTTAGAAAGGCT 1480
|||||

Db 1370 AGCGCGGCGAAGGTAAANACGGCATTTCAATTTAGCAANAGAAACCTCTTTAGAAAGGCT 1429
QY 1481 CAACCATCAATGATATCAGCGCAAGAAAGCGCGCATTTGTTGTTGGCGGATATTG 1540
Db 1430 CAACCATCAATGATATCAGCGCAAGAAAGCGCGCATTTGTTGTTGGCGGATATTG 1489
QY 1541 CGTTAATTTGACGGCAATTAACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTT 1600
Db 1490 CGTTAATTTGACGGCAATTAACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTT 1549
QY 1601 TTGTGGAGACATCGGGCATTTATTATCCATTGACAGCAATGCAATTTGTTAAACAAAG 1660
Db 1550 TTGTGGAGACATCGGGCGATGATTTATCTCAAGCAATGCAATTTGTTGACGCAAG 1609
QY 1661 AGTGGTTGTAGACCTGATGATTAACAAATTTGAAGCGAAGACCCCTTGGCAATATA 1720
Db 1610 AGTGGTTGTAGACCGGATATGATTTAATGAGAAACAGCAGCAGCAATA 1669
QY 1721 CCGGTATTAATGATCAATTTCCCAACAGGCGGTTGAAGCAAGCGACCTTAANAATA 1780
Db 1670 CTTCAAGACCATGATATACACGGGATCGGGAATAGTGGCAGCACCCTCAAGCAACA 1729
QY 1781 CGCAACTCAAAAACACGCTAACCAATACAACTATTTCAAAATTTCTGAAAAACGCTGGA 1840
Db 1730 AAGAA--AAGACAACATTTAACAAACACAACTCTTTGAGAGTATATCTAAAAAAGTACCT 1786
QY 1841 CAATGAATTAACGGCATCAAGAAACTTTACCGTTAATAGCTCAATCAACATCGGAAGCA 1900
Db 1787 TTGTTAATCATCTCTCTAATCAACGCATCTATGTCATATAGCTCCATTAATTT--ATCCA 1843
QY 1901 ACTCCACATTAATCTCCATAGTAAAGGTACAGCTGGCGGAGGCTTCAGATTGATGGAG 1960
Db 1844 ATGCGACCTTAATCTTTGGAGTGAGGCTCGGAGCGGTGGCGGCTTGAGATTACAAACG 1903
QY 1961 ATATTAC-----TTCTAAAGCGGGAATTTAACCATTTATCTGCGCGATGGG 2008
Db 1904 ATATTACCCCGGTGATGATACAGAGGTGCAAACTTAACAAATTTACTCAGCGCGCTGG 1963
QY 2009 TTGATGTTCAATAAATAATTAACGCTTGATCAGGCTTTTAAATATATACCGCGCTCCG 2068
Db 1964 TTGATGTTCAATAAATAATCTCCTCGGCGCAGGTAAACATAACATTTACAGCTAAC 2023
QY 2069 TAGCTTTGAGGTGAAATTAACAAAGCAGCGCGGCAATGCTAAATTTGTCGCC 2128
Db 2024 AAGATATCGCTTTGAGAAAGAGCAACAACTC-----ATTACAGGTC 2068
QY 2129 AGGCACTGTAAACATTAACAGAGAGGAAAGATTTTCAGGCTTAACAAGTATCTTTAA 2188
Db 2069 AAGGACTATTTACC--TCAGGCAATCAAAAAGGTTTTAGATTTTAATATGCTCTCTAA 2125
QY 2189 ACGGAACGGGTAAAGCTCTGAATATTCATCTCATGTAATTAATTTA-----ACCC 2239
Db 2126 ACGCACTGGCAGCGGCTGCAATTCACCACATAAAGAAACCAATAATACGCTATCAAA 2185
QY 2240 ACAATCTTAGTGCAATTAACATATCTGGGAATTAACAAATTAACAACTA---CGA 2296
Db 2186 ATAAATTTGAAGGACTTTAAATATTTTCAGGGAAGTGAACATCTCAATGGTTTTACCTA 2245
QY 2297 GAAAGAACCTCTGATTTGGCAACACAGCATGATTCGCATCGCAACGCTGCTCTTA 2356
Db 2246 AAAATGAAGTGGATATGATTAATTCNAGGACGCACTTACTGGAATTTAACTCGAAG 2305
QY 2357 ATCTAGAGACGCGCAATTTTACCTT---TATTAAATACATTTCAAGCAATAGCAAG 2413
Db 2306 TGGATATGATAAATTCAAAGGACGCTCTACTATTGACTCCAGAGGAGCGATAGTGAG 2365
QY 2414 GCTTAACACACAGTATAGAGCTCTCAGGCGTGAATTTTAACGCGTAAATGCAACA 2473
Db 2366 GCACACTTACCCAGCGCTTATTAATTTAAACGGTATATCATTCACCAAGACACT---ACCT 2422
QY 2474 TGTCAATTCATCTCAAGAGGCGGAAAGTTAATTTTCAAAATTAACCAACCAAGCAACA 2533
Db 2423 TTAATGTTGAAGCAATCAAGAGTCAACTTTGACATCAAGCACCNAATAGGATAAATA 2482

Db 4628 TGAATACATTCACCGGTATAGCAAGCGTAGATGAAGTAATGAAGCGAAGCATCC 4687
 QY 4559 TTGAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTAAGTAAACTTGGTGTA 4618
 Db 4688 TTGAGAAGGTAAGATTTATCTGATGAAGAAAGAAAGCGCTAGCTAAACTTGGCGTAA 4747
 QY 4619 GTGCTGACGTTTGTGAGCCAAATATACAAATTAACAGTCAATACAAATGAATTTA 4678
 Db 4748 GTGCTGACGTTTATTTAGCCAAATATACAAATTAACAGTCAATACAAATGAATTTG 4807
 QY 4679 CAACACAGCGTCAAGTCAAGTGAATATTTCTCAAGGTAAGCGGTGTTCTCAAGTGTA 4738
 Db 4808 CAACACAGCGTCAAGTCAAGTGAATATTTCTCAAGGTAAGCGGTGTTCTCAACACAGT 4867
 QY 4739 ATGGCGCAGCAGTATGTACCAATGTGCTGACGATGGACGCGGTAGTCAAGTAATGACA 4798
 Db 4868 ATGGCGCAGCGGTGCGGTTAATATCTGATATACGCGGCTAGTCAAGTAATGACA 4927
 QY 4799 AGGTAGATTTATCCCTGCAATCAAGTCAATTTATTTCTGATATTTACTGTGGGTTA 4858
 Db 4928 AGGTAGATTTATCCCTGCAATCAAGTCAATTTATTTCTGATATTTACTGTGGGTTA 4987
 QY 4859 AAGTTCAGTACGGCTTTACCCATCTTGTAAAAAATACGAGAAATACAAATGAATTT 4918
 Db 4988 AAGTTCAGTACGGCTTTACCCATCTTGTAAAAAATACGAGAAATACAAATGAATTT 5047
 QY 4919 TTAACAGGTTATTTATG 4937
 Db 5048 TTAACAGGTTATTTATG 5066

RESULT 14

AAT90993

ID AAT90993 standard; DNA; 4803 BP.

AC AAT90993;

XX AAT90993;

XX 14-APR-1998 (first entry)

XX Non-typeable Haemophilus high mol.wt. surface protein HMW4 gene.

XX Non-typeable Haemophilus; high molecular weight surface protein;

XX HMW4; immunogen; vaccine; otitis media; ss.

XX Haemophilus influenzae strain 5.

XX Key Location/Qualifiers

FT CDS 1..4803

FT /*tag= a

FT /transl_except= (pos:1114..1116, aa:Thr)

FT /transl_except= (pos:1198..1200, aa:Asp)

FT

XX W09736914-A1.

XX 09-OCT-1997.

XX 01-APR-1997; 97W0-US04707.

XX 01-APR-1996; 96US-0617697.

XX (BARE/) BARENKAMP S J.

XX Barenkamp SJ;

XX WPI; 1997-503038/46.

XX P-PSDB; AAW30292.

XX High molecular weight proteins of non-typeable Haemophilus

XX influenzae - useful for vaccine production

XX Claim 1; Page 90-92; 183pp; English.

XX

CC 2 Novel nucleic acids (AAT90992 and AAT90993) respectively comprise the
 CC coding sequences of high molecular weight surface proteins HMW3
 CC (see AAW30291) and HMW4 (see AAW30292) of Haemophilus influenzae strain
 CC 5. HMW3 has a mol.wt. of 125 kDa and HMW4 has a mol.wt. of 123
 CC kDa. Both are antigenically related to HMW1 (see AAW30293) and HMW2
 CC (see AAW30294) obtained from non-typeable H. influenzae strain 12.
 CC Vectors containing HMW nucleic acids for transformation of a host
 CC are claimed. HMW nucleic acids can also be used as probes for the
 CC detection of related genes. HMW proteins, conjugates and peptides
 CC can be used in vaccines, as immunogens for preparation of
 CC antibodies and as antigens for detection of these antibodies.
 XX

SQ Sequence 4803 BP; 1714 A; 928 C; 971 G; 1190 T; 0 other;

Query Match 45.5%; Score 2245; DB 18; Length 4803;

Best Local Similarity 70.6%; Pred. No. 0;

Matches 3411; Conservative 0; Mismatches 990; Indels 432; Gaps 14;

QY 352 ATGACAAAGATATATCGTCTCAAAATTCAGCAACGCGCTGAATGTTGGTGGTGTCT 411
 Db 1 ATGACAAAGATATATCGTCTCAAAATTCAGCAACGCGCTGAATGTTGGTGGTGTCT 60
 QY 412 CAATTGGCAGCGGGTGTGACCATTTCCACAGAAAAGGCTTCCGCTATGTTACTCTTT 471
 Db 61 GAATTGACACGCGGGTGTGACCATTTCCACAGAAAAGGCGTGAATAAACCCTGTCGTACG 120
 QY 472 AGGTGTAACACACTAGCGTTAAAGCCACTTTCGCTATGTTACTATCTTTAGTGTAACA 531
 Db 121 AAGTAGCCCACTTGGCGTTAAAGCCACTTTCGCTATGTTACTATCTTTAGTGTAACA 180
 QY 532 TCTATTCACAACTCTGTTTTAGCAAGCGGCTTACAAAGGAATGGATGTAGTACACGGCACA 591
 Db 181 TCCATTCCGCAATCTGTTTTAGCGAGCGGTTTACAGGGAATGAGCGCTGCTACACGGTACA 240
 QY 592 GCCACTATGCAAGTAGTGGTAATAAAACCATTATCCGCAACAGCTGTGACGCTATCAATT 651
 Db 241 GCAACCATGCAAGTAGAGCGCAATAAACCATTATCCGTAATAGCGTCAATGCTATCATC 300
 QY 652 AATTGGAACAATTTAAACATCGACCAAAATGAAATGGTGCAGTGTTCACAGAAAACAAAC 711
 Db 301 AATTGGAACAATTTAAACATTTGACCAAAATGAAATGGTGCAGTGTTCACAGAAAACAAAC 360
 QY 712 AACTCCCGCGGTATTCACACCGTGTACATCTTAACAAATCTCCCAATTAATAAGGGATTTTA 771
 Db 361 AACTCTGCCGTTTCAACCGTGTACATCTGACCAAAATCTCCCAATTAATAAGGGATTTTA 420
 QY 772 GATTCTAACGGCAAGTCTTTTTTAATCAACCCAAATGTTATCATATAGTAAAGACGCA 831
 Db 421 GATTCTAACGGCAAGTCTTTTTTAATCAACCCAAATGTTATCATATAGTAAAGACGCA 480
 QY 832 ATTTAAACACTAATGGCTTTTACGGCTTCTAGCGTAGACATTTCTAACGAAAACATCAAG 891
 Db 481 ATTTAAACACTAATGGCTTTTACGGCTTCTAGCGTAGACATTTCTAACGAAAACATCAAG 540
 QY 892 GCGGTAAATTTCACTTCGACCAAAACCAAGATAAAGCGCTCGCTGAAATGTTGAATCAC 951
 Db 541 GCGGTAAATTTCACTTCGACCAAAACCAAGATAAAGCGCTCGCTGAAATGTTGAATCAC 600
 QY 952 GGTTTAAATTTACTCTCGGTAAAGACGCGAGTGTAAATCTTATTTGGTGCAGAACTGAAAAAC 1011
 Db 601 GGTTTAAATTTACTCTCGGTAAAGACGCGAGTGTAAATCTTATTTGGTGCAGAACTGAAAAAC 660
 QY 1012 GAGGGTGTGATTTAGCGTAAATGGTGGCAGCATTTCTTTACTTCGCGAGGCAAAAATCACC 1071
 Db 661 GAGGGTGTGATTTAGCGTAAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 720
 QY 1072 ATCAGCGATATAATAAACCCCAACCATTTACTTACAGCATTTGCGCGGCTCGGAAATGAAGCG 1131
 Db 721 ATCAGCGATATAATAAACCCCAACCATTTACTTACAGCATTTGCGCGGCTCGGAAATGAAGCG 780
 QY 1132 GTCAATCTGGCGGATATTTTGGCAAGGCGGTAAACATTAATGTCGCTGCTGCCACTATT 1191
 Db 781 ATCAATCTGGCGGATATTTTGGCAAGGCGGTAAACATTAATGTCGCTGCTGCCACTATT 840

D	b	4051	GCAAGCACTGGTGATTTAACTATTGGAATACTGCAAAAGTTGAAGCGAAAAATGGAGCT	4110
Q	y	4090	GCAACCTTAACCGCAACAGGGAAATACCTTGACTACTGAAGCGGTCTTACGATCACTTCA	4149
D	b	4111	GCAACCTTAACTGCTGNATCAGGCCAAATTAACCAACCAACAGGCCTTAGCATTAACCTCA	4170
Q	y	4150	ACTTAGGGTCAGGTAGACCTCTTGGCTCGAATGGTAGCATCGCAGGAGCAATTAATGGCT	4209
D	b	4171	ACCAATGGTCAGACAACTCTTACAGCCAGGATAGCAGTATCCGAGGAACAACTTAATGGCT	4230
Q	y	4210	GCTAATGTGACATTAATACTACAGGCACCTTAACCAACCCGTCGAGGCTCGGATATTAA	4269
D	b	4231	GCTAATGTGACGTAAATACCAACAGGCACCTTTAACTACTACAGGGGATTCAAAGATTAAC	4290
Q	y	4270	GCAACCAAGCGGCACCTCTGGTTATTACGCAAAAGATGCTTAAGCTAAATGGTGTGATCATCA	4329
D	b	4291	GCAACCAAGTGTTACCTTAACCACTCAATATGCAAAAGATGCCAAATTAGATGGTGTGCTCATCA	4350
Q	y	4330	GGTGATGTACAGAGTAGTGCAATGTCAGTCAACCAACGCGGCTCTGGTGTGTGCTCGCGCA	4389
D	b	4351	GGTCAGCCGACAGTAGTAATATGCAACTAACCAAGTGGCTCTGGTACGTGACTCGGAAA	4410
Q	y	4390	ACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACACAGTAATTGGGTTAAATATCAAT	4449
D	b	4411	ACCTCAAGCAGCGTGAATATCACCGGGGATTTAAACACAAATGAATGGGTTAAATATCATTT	4470
Q	y	4450	TCGAAAGATGGTGAACAACTGTGCGCTTAGAGGCAAGGAAATTTGAGGTGAAATATATC	4509
D	b	4471	TCGGAATAATGGTAGAACACTGTGCGCTTAGAGGCAAGGAAATTTGATGAAATATATC	4530
Q	y	4510	CAGCCAGGTGTAGCAAGTGTAGAGAAGTAATTTGAAGCGAAACCGCTCTCGAAAAAGTA	4569
D	b	4531	CAACCAGGTGTAGCAAGCGTAGAAGAGTAATTTGAAGCGAAACCGCTCTCGAAGAGGTA	4590
Q	y	4570	AAAGATTTATCTGATGAAGAAGAAACAAATAGCTAAACTTGGCTTAAGTCTGTGACGT	4629
D	b	4591	AAAGATTTATCTGATGAAGAAGAAACAAATAGCTAAACTTGGCTTAAGTCTGTGACGT	4650
Q	y	4630	TTTGTGTAGCGCAATAATACAATACAGTCAATACACAAATGAATTTACACCAACGCG	4689
D	b	4651	TTCTGTGTAGCGCAATAATGCCATTACGGTTTAATACAAAAGAGGTTTACAAACCAACCA	4710
Q	y	4690	TCAAGTCAAGTGATTAATTTCTGAGGTAGAGGTAGGCGGTGTTCTTCAAGTGGTAAATGGCGCAGCA	4749
D	b	4711	TCAAGTCAAGTGACAATTTCTGAGGTAGAGGTAGGCGGTGTTCTTCAAGTGGTAAATGGCGCAGCA	4770
Q	y	4750	GTATGTACCAANTTGTGTGACGATGGACAGCG	4782
D	b	4771	GTATGTACCAANTTGTGTGACGATGGACAGCAG	4803

RESOL 13
AAO49517

AAQ49511:

26-APR-1994 (first entry)

sequence encoding high molecular weight protein 4 (HMW4).

S Haemophilus influenzae

Key	Location/Qualifiers
CDs	3 450

product=

```

T
/*tag= b
/transl ev

```

```

, trans_except= TCT encodes Threonine.

```


1607 TCTATTACGGGAGGGCAAGGTTTAAAGTTTATTGCAAAATCAAAATTAATTTCACTCAT 1666
1667 AAATTTGATGGCGAAATTAACATATCTCGAATAGTAAATTAACCAACACGCAAAAA 1726
2302 AACACCTCGTATTGGCAAAACGAGCCATGATTCGCACTGGAAGCTGAGTCTTAATCTA 2361
1727 GATGTTAAATCTGGNATGCAATCAAAAGACTCTTACTGGAATGTTTCTTCTTACTTTG 1786
2362 GAGACAGGCGCAATTTTACCTTTTAAATACATTTCAAG-----AATAGCAAGGC 2415
1787 AATACGGTGCAAAATTTTACCTTTTATAAAATTCGTTGATAGCGGCTCAAAATTCGCCAAGT 1846
2416 TTAACAACACAGTATAGAGCTCTGAGGGGTGAATTTTACGCGGTAAATGCGCAATG 2475
1847 TTGAGGTCATCAGGTAGAGTTTTCAGGGGTACATTTTAAAGGTCATCGGAGCAAAACA 1906
2476 TCATTCAATCTCAAGAGGAGCGAAGTTAATTTCAAATTTAAACCAACAGAGAACATG 2535
1907 AACTTCAACATCGAGCTAACGCAAAAGCCTTATTTAAATTTAAACCAACAGCGCTACA 1966
2536 AACACAGCAACCTTTACCAATTCGGTTTTTACCAATATCACAGCCACTG-----GT 2589
1967 GACCAAAAGAAATTTACCTTTTAAAGGCAACATTTACGCTACCGGTAAACAGT 2026
2590 GGGGCTCTGTTTTTTTATATATATGCAACCATTTCTGCGAGAGGGCTGAGTTAAAA 2649
2027 GATAGCTCTGATGTTGACATACAGCCCATCTTACCTAGAGCTGCCGGCATTAAC 2086
2650 ATGAGTGAATTAATATCTTAACGGGGCTAATTTTACCTTTAAATTTCCCATGTTCCGGGC 2709
2087 ATGATTTCAATTAACATTTACCGGGGCTTGACTTTTCCATTAACATCCCATTAATCGCAAT 2146
2710 GATGACCTTTTAAATCAACAAAGACTTAACCATTAATTAACCAATTCAAATTTGACG 2769
2147 AGTAATGCTTTTGAATCAAAAGACTTAACATTAATTAATGCACTGGCTCGAATTTTANGT 2206
2770 CTCAGACAGCAAGATGATTTTATGAGGGTACGCAAGCAATGCCATCAATTAATCAAC 2829
2207 CTTAAGCAACAGAGATTCCTTTTATATGAATACAGCAAAACAGCCATTAATCAAGT 2266
2830 TACAACATATCAATCTGGGGGTAAATGTCACCCCTGGTGGCAAAACTCAAGCAGCAGC 2889
2267 CATATCTAACCATCTTGGGGCAATGTCACCTGAGTGGGGAATTTCAAGCAGTAGC 2326
2890 ATTACGGGAATATTACTACGAGAAAGCAGCAATGTTACGCTAGAGCCCAATTAACGCC 2949
2327 ATTACGGCAATATCANATCACCAATTAAGCAATGTTACATTTACAAGCTGACACCAGC 2386
2950 CTAATCAGCAAAACATTAAGGATAGAGTTTATAAACTTTGGCAGCTTGCTCGTTAATGGG 3009
2387 AACAGCAACAGGCTTGAAGAAAGAACTTAACCTTTGGCAATATATCTGTTGAGGG 2446
3010 AGTTTAATGTTTAACTGGGCAATGAGATATTAAGGCAATCTCACTATTTCAGAAAGC 3069
2447 AATTTAAGCTAATCTGTTGCAATGCAACATGTCGCAATCTTCTTATTCGAAGAT 2506
3070 GCCACTTTTAAAGAAAGACTAGAGTACCCTTAATATCACCGCAATTTTACCATAAT 3129
2507 TCCACATTTTAAAGGAGAGCCAGTGACAACCTTAACATCACCGCACCTTTTACCAACAAC 2566
3130 GGCACCTCCCAATTAATTAACACAGAGGTGGTAAACCT---TGGCAATGTTACCAAT 3186
2567 GGTACCGCCCAACATTAATATAAAACAAGAGGTGGTAAACCTTCCAGGCGATATATCAAT 2626
3187 GATGCTGATTTAAACATTAACCACTACGCTTAAACGCAACCAAGAGCATCATCGGCGGA 3246
2627 AAGAGTGGTTAATATCACTACTACGCTTACGCTTACGCTTACGCTTACGCTTACGCT 2686
3247 GATTAATCAACAAAAAGGAGCTTAAATATATACAGACAGTAAATGATGCTGAAATC 3306

2687 AATATAACTAACGAAAAAGGCGACTTTAAACATCAAGAATATTAAAGCCGACGCCGAAATC 2746
3307 CAATTTGGCGGCAATATCTCCGAAAAAGAGCAACCTCAGCATTTCTTCGATTAATTT 3366
2747 CAATTTGGCGGCAATATCTCACAAGAAAGCAATCTCACAATTTCTTCTGATTAAGTA 2806
3367 AATATCACCACCAACACATCAAAATCAAAAGGGTATTGATGAGAGGACTCTAGTTTCAGAT 3426
2807 AATATTACCAATCAGATCAAAATCAAAAGCGGCTTGAAGGGGGGCTTCTGATTTCAAGT 2866
3427 GCGACAGTAATGCCAAGCTTAACCTTATTAACCAAGAAATTTGAATTTGACAGAGACCTA 3486
2867 GAGCGAAAAATGCTTAACCTTAACCTTATTAACCAAGAGTTTAAATTTGCGAGGAGACCTA 2926
3487 AGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCCAAGATGTCAGAGATTTAACATTT 3546
2927 AATATTTCAGGCTTTAATTAAGCAGAAATTAACGCTAATAATGGCAGTGTATTACTATT 2986
3547 GCGAACAGTAATGACGGTAAACGCGTCCGAGGCAACCAAGCACTAATTTTAAACATGTT 3606
2987 GGCAATGCTAGCGGTGGTAA-----TGCTGATGCTAAAAAGTGACTTTTGACAAGGTT 3040
3607 AAGATTTCAAAATCTCTGCTGAGCGTCACATGTCACACTTAATAGCAAACTGAAACACA 3666
3041 AAGATTTCAAAATCTCGACTGAGCGTCACATGTAACACTTAATAGCGAAGTGAACAGC 3100
3667 TCTAGCAGCAATGGCGGACGCTGAAAGCAATAGCAGCAACGATACCGGCTTAATCTATTACT 3726
3101 TCT-----AATGGTAGTAGCAATGCTGGTAAATGTAACAGCACCGGTTTAACCATTTCC 3154
3727 GCAAAATGTTAGAATTAACAAAGATATTACTTCTCAAAACAGTAAATATCAACGCG- 3785
3155 GCAAAAGATGTAACGGTAAACATTAACGTTACCTCCCAAGACAAATAATATCTCTGCC 3214
3786 ----- 3785
3215 GCAGCAGGAATGTAACCAACCAAGAGCAACTATCAATGCAACACAGGCGAGCGTG 3274
3786 ----- 3785
3275 GAAGTAAGTCTCAAAATGGTACAAATTAAGAGCAACATTAACCTCGCAAAATGTAACAGTG 3334
3786 ----- 3785
3335 ACAGCAACAGAAATCTTGTGTACACAGAGAATGCTGTCATTAATGCAACACAGGCGGACA 3394
3786 ----- 3785
3395 GTAAACATTTAGTACAAAAACAGGGGATATTAAAGGTGGAATTTGAATCAACTTCCGGTAAT 3454
3786 ----- 3785
3455 GTAAATATTACAGCGAGCGCAATACACTTAAGGTAAAGTAAATATCACTGGTCAAGATGTA 3514
3786 -----GTCCGAAAAAGTTTACCACACAGCAGGCTCGACCAATTAACGCAACA 3831
3515 ACAGTAACAGCGGATGACAGGCGCTTGACAACTACAGCAGGCTCAACCATTAAGTCCGACA 3574
3832 AATGGCAAGCAAGTATTACAAACCAACAGAGGTGATATCAGCGGTACG----- 3879
3575 ACAGGCAATGCAAAATATTACAAACCAACAGAGGTGATATCAACGCTAAAGTTGAATCCAGC 3634
3880 -----ATTTCCGGT 3888
3635 TCCGGCTCTGTAAACACTTTGTAACCTGGAGCAACTCTGCTGTAGGTAAATATTTCAGGT 3694
3889 AACACGTAAGTGTTAGCGGCACT---GGTGAATTTAAACCACTAAATCCGCTCAAAAATTT 3945
3695 AACACTGTTACTATTACTGCGGATAGCGGTAAATTTAACTTCCACAGTAGTCTTACAAAT 3754
3946 GAAGCGAAATCGGCTGAGGCTTAATGTAACAAGTGCACAGGTACAATTTGGCGGTACAATTT 4005
3755 AATGGGACTAATAGTGTAAACCA-----CCTCAAGCAATCAGCGGATATTGAAGGTACAATTT 3811

us-10-092-880-3.rng

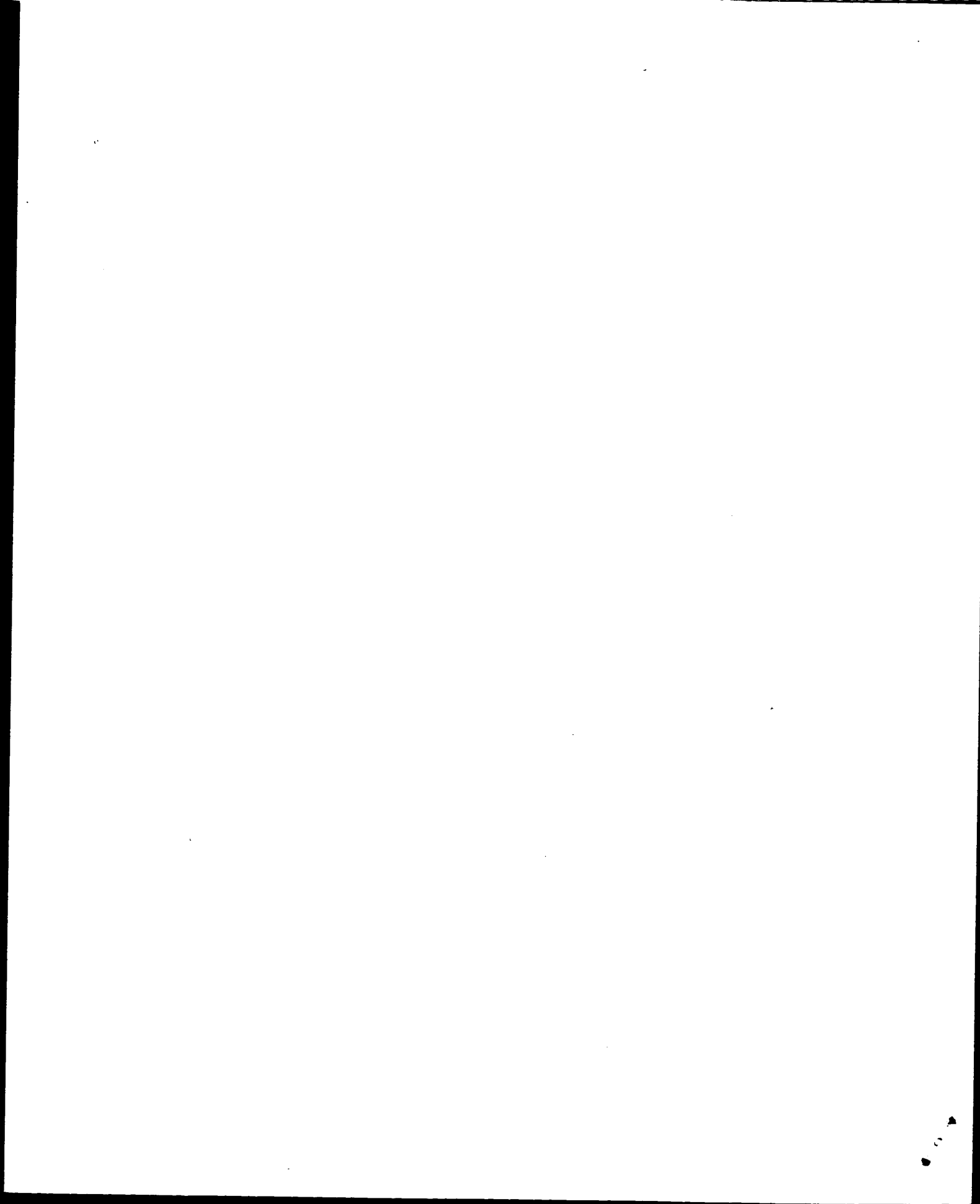
Tue Mar 25 08:39:01 2003

```

QY 4006 TCCGGTAATACGGTAATAATGTTACGGCAAAACGGTGGCGGATTTAACAGTTGGGAATGGCGCA 4065
Db 3812 TCTGGTAATACAGTAATGTTAGCAAGCACTGGTGTAACTATTGGAATAGTGCA 3871
QY 4066 GAAATTAATGCGACAGAGAGCTGCAACCTTAAACCGCAACAGGGAAATACCTTGACTACT 4125
Db 3872 AAAGTTCAAGCGAAATGAGCTGCAACCTTAACTGCAATCAGGCAATTAACCC 3931
QY 4126 GAAGCCGGTTCTAGCTACATCTCAACTAAGGTCAGGTAGACCTCTTGCTCAGATGGT 4185
Db 3932 CAACAGGCTTAGCTATCTCACTCAAGCAATGGTCAGACACCTCTACACCGCAAGATAGC 3991
QY 4186 AGCATCGCAGGAAGCATTAATGCTGTATGTACATTAATTAATACAGGACCTTAACG 4245
Db 3992 AGTATCGCAGGAACATTAATGCTGTATGTACATTAATTAATACAGGACCTTTAACT 4051
QY 4246 ACCGTGCGAGCTCGGATTAATTAAGCAACGAGCGGACCTTGGTTATTAACGCAAGAT 4305
Db 4052 ACTACAGGGATTCAGAGATTAACGCAACCGTGGTACCTTAACATCAATGCAAAAGAT 4111
QY 4306 GCTAAGCTAATGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4365
Db 4112 GCCAAATTAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4171
QY 4366 GCTTCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 4425
Db 4172 GGTCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 4231
QY 4426 ACAGTAATGGTTAAATATATATATATATATATATATATATATATATATATATATATATAT 4485
Db 4232 ACATAATGGTTAAATATATATATATATATATATATATATATATATATATATATATATAT 4291
QY 4486 AAGCAATGGTTAAATATATATATATATATATATATATATATATATATATATATATATATAT 4545
Db 4292 AAGCAATGGTTAAATATATATATATATATATATATATATATATATATATATATATATAT 4351
QY 4546 CGCAACCGCTCTTGAGAGGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAGCT 4605
Db 4352 CGCAACCGCTCTTGAGAGGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAGCT 4411
QY 4606 AAATCTGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 4665
Db 4412 AAATCTGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 4471
QY 4666 CAATATGAATTTACACACGAGCTGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4725
Db 4472 CAATATGAATTTACACACGAGCTGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4531
QY 4726 TTCTCAAGTGTAAATGGGCGACGAGTATGTACCAATGTTGCTGACGATGGACAGCCGTAG 4785
Db 4532 TTCTCAAGTGTAAATGGGCGACGAGTATGTACCAATGTTGCTGACGATGGACAGCCGTAG 4591
QY 4786 TCAGTAATTCACAAGGTAGATTTTCATCCTGCAATGAAGTCAATTTATTTTCGTATATTT 4845
Db 4592 TCAGTAATTCACAAGGTAGATTTTCATCCTGCAATGAAGTCAATTTATTTTCGTATATTT 4651
QY 4846 ACTGTGTGGTTAAAGTTTACGTACGGGCTTTACCCATCTGTGAAAAATTA 4896
Db 4652 ACTGTGTGGTTAAAGTTTACGTACGGGCTTTACCCATCTGTGAAAAATTA 4702

```

Search completed: March 21, 2003, 21:15:19
Job time : 1049 secs



QY 61 CACCTTTTTCAGCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATA 120
DB 61 CACCTTTTTCAGCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATA 120
QY 121 AATGGTAATATCTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
DB 121 AATGGTAATATCTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
DB 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
DB 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 301 GAAGCCAAATGATAAAGTAATTTAATTTGTTCAACTTAACCTTAGGAGAAAATATGAACAAG 360
DB 301 GAAGCCAAATGATAAAGTAATTTAATTTGTTCAACTTAACCTTAGGAGAAAATATGAACAAG 360
QY 361 ATATATCGTCTCAAAATCAGCAACCGCTGAATGCTTTGTTGCTGTGTGCTGTGCTGTGCTGTG 420
DB 361 ATATATCGTCTCAAAATCAGCAACCGCTGAATGCTTTGTTGCTGTGTGCTGTGCTGTGCTGTG 420
QY 421 CGGGGTTGTGACCATTCACAGAAAAGGCTTCGGCTATCTTACTATCTTTAGGTGTGAAC 480
DB 421 CGGGGTTGTGACCATTCACAGAAAAGGCTTCGGCTATCTTACTATCTTTAGGTGTGAAC 480
QY 481 CACTTAGCTTAAGACCATTTCCGCTATGTTACTATCTTTAGGTGTGAACATCTATTTCCA 540
DB 481 CACTTAGCTTAAGACCATTTCCGCTATGTTACTATCTTTAGGTGTGAACATCTATTTCCA 540
QY 541 CAATCTGTTTGAACGCGCTTACAAGGAATGATGTAGTACAGGACACAGCCACTATG 600
DB 541 CAATCTGTTTGAACGCGCTTACAAGGAATGATGTAGTACAGGACACAGCCACTATG 600
QY 601 CAAGTAGTGTATATAAACCATTATCCGCAACAGTGTGAGCTATCATATTAATTTGAAA 660
DB 601 CAAGTAGTGTATATAAACCATTATCCGCAACAGTGTGAGCTATCATATTAATTTGAAA 660
QY 661 CAATTTAATCATGACCAAAATGAAATGTTGCGAGTTTACAAGAAAACAACTCCGCC 720
DB 661 CAATTTAATCATGACCAAAATGAAATGTTGCGAGTTTACAAGAAAACAACTCCGCC 720
QY 721 GTATTCACCGCTGTACATCTAACCAATCTCCCAATTAAGAGGATTTAGATTTCTAAC 780
DB 721 GTATTCACCGCTGTACATCTAACCAATCTCCCAATTAAGAGGATTTAGATTTCTAAC 780
QY 781 GGACAAAGTCTTTTAAATCAACCAAAATGGTATCACAATAGTAAAGACGCAATTTAATAC 840
DB 781 GGACAAAGTCTTTTAAATCAACCAAAATGGTATCACAATAGTAAAGACGCAATTTAATAC 840
QY 841 ACTAATGGCTTTACGGCTTACGCTAGACATTTCAAGCAAAATCAAGCGCGTAAAT 900
DB 841 ACTAATGGCTTTACGGCTTACGCTAGACATTTCAAGCAAAATCAAGCGCGTAAAT 900
QY 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGAATACGGTTTAAAT 960
DB 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGAATACGGTTTAAAT 960
QY 961 ACTGTGCTGTAAGACGCGAGTGAATCTTTATTTGTTGCAAGTGAAGAACGAGGCTGTG 1020
DB 961 ACTGTGCTGTAAGACGCGAGTGAATCTTTATTTGTTGCAAGTGAAGAACGAGGCTGTG 1020
QY 1021 ATTACCGTAAATGGTGGGAGCATTTCTTTACTCGAGGCAAAAAATCACCATCAGCAT 1080
DB 1021 ATTACCGTAAATGGTGGGAGCATTTCTTTTACTCGAGGCAAAAAATCACCATCAGCAT 1080
QY 1081 ATAATAAACCCCAACCATTTACTTACAGCATTTCCGCGCTGAAATGAAGCGTCAATCTG 1140
DB 1081 ATAATAAACCCCAACCATTTACTTACAGCATTTCCGCGCTGAAATGAAGCGTCAATCTG 1140

QY 1141 GCGGATATTTTTCGCAAGGCGGTAACTAATATGTCGCTGCTGCCACTATTTTGAACCAA 1200
DB 1141 GCGGATATTTTTCGCAAGGCGGTAACTAATATGTCGCTGCTGCCACTATTTTGAACCAA 1200
QY 1201 GGTAAACTTTTCTGCTGATTTCTGTAAGCAAGATAAAGCGGCAATATTTGTTCTTCGCGC 1260
DB 1201 GGTAAACTTTTCTGCTGATTTCTGTAAGCAAGATAAAGCGGCAATATTTGTTCTTCGCGC 1260
QY 1261 AAAGAGGTTGAAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
DB 1261 AAAGAGGTTGAAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
QY 1321 GGCAAGCTGATGATTTACAGCGGATAAAGTCAATTTAAACAGGTGCGATTTATCGACCTT 1380
DB 1321 GGCAAGCTGATGATTTACAGCGGATAAAGTCAATTTAAACAGGTGCGATTTATCGACCTT 1380
QY 1381 TCAGGTAAAGAGGCGGAGAACTTACCTTGGCGGTGACGAGCGCGGCAAGGTAAAGAC 1440
DB 1381 TCAGGTAAAGAGGCGGAGAACTTACCTTGGCGGTGACGAGCGCGGCAAGGTAAAGAC 1440
QY 1441 GGCATTCATTTAGCAAAAGAAACCTTTTACAAAAAGGCTCAACCATCAATGTATCAGC 1500
DB 1441 GGCATTCATTTAGCAAAAGAAACCTTTTACAAAAAGGCTCAACCATCAATGTATCAGC 1500
QY 1501 AAAGAAAAAGGCGGACCGCTATTTGTGGGGGATATTTGCGTTAAATGACGCGCAATTT 1560
DB 1501 AAAGAAAAAGGCGGACCGCTATTTGTGGGGGATATTTGCGTTAAATGACGCGCAATTT 1560
QY 1561 AACGCTCAAGGTAGTGTGATATCGTAAAAACCGGTGTTTGTGGAGACATCGGGCAT 1620
DB 1561 AACGCTCAAGGTAGTGTGATATCGTAAAAACCGGTGTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTTATCCATTTGACGCAATGCAATTTGTTAAACAAAAAGAGTGTGTAGACCCCTGAT 1680
DB 1621 TATTTATCCATTTGACGCAATGCAATTTGTTAAACAAAAAGAGTGTGTAGACCCCTGAT 1680
QY 1681 GATGTAACTAATGAAGCGGAGACCCCTTCGCAATTAATACCGGTATATAATGATGATTC 1740
DB 1681 GATGTAACTAATGAAGCGGAGACCCCTTCGCAATTAATACCGGTATATAATGATGATTC 1740
QY 1741 CCAACAGGACCGGTGAAGCAAGCGCCCTTAAAAAAATAGGCACTCAAAACAGCTA 1800
DB 1741 CCAACAGGACCGGTGAAGCAAGCGCCCTTAAAAAAATAGGCACTCAAAACAGCTA 1800
QY 1801 ACCAATCAACTATTTCAAAATTTATCTGAAAAACGCGCTGGACAATGAATATAACGGCATCA 1860
DB 1801 ACCAATCAACTATTTCAAAATTTATCTGAAAAACGCGCTGGACAATGAATATAACGGCATCA 1860
QY 1861 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCACCTTAATTTCTCCAT 1920
DB 1861 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCACCTTAATTTCTCCAT 1920
QY 1921 AGTAAAGGTGAGGTTGCGGAGGCTTCAGATTTGATGAGATATTTACTTCTAAAGCGGA 1980
DB 1921 AGTAAAGGTGAGGTTGCGGAGGCTTCAGATTTGATGAGATATTTACTTCTAAAGCGGA 1980
QY 1981 AATTTAACTATTTCTTTCGCGGATGGTTGATGTTTCAATAAATATAGCTGTATCAG 2040
DB 1981 AATTTAACTATTTCTTTCGCGGATGGTTGATGTTTCAATAAATATAGCTGTATCAG 2040
QY 2041 GGTTTTTTAAATTTACCGCGCTTCGTTAGCTTTTGAAGGTGGAATACAAACGACGC 2100
DB 2041 GGTTTTTTAAATTTACCGCGCTTCGTTAGCTTTTGAAGGTGGAATACAAACGACGC 2100
QY 2101 GACCGGCAAAATGCTAAATTTGTCGCCAGGCACTGTAACCATACAGGAGGAGGAAA 2160
DB 2101 GACCGGCAAAATGCTAAATTTGTCGCCAGGCACTGTAACCATACAGGAGGAGGAAA 2160
QY 2161 GATTTAGGCTTACCAAGCTATCTTTAAACGGAAGCGGTAAAGCTCTGAATATCATTTCA 2220
DB 2161 GATTTAGGCTTACCAAGCTATCTTTAAACGGAAGCGGTAAAGCTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATATTTTAAACCCCAATTTAGTGGCACAATTTACATATCTCGGGAATATA 2280

2221 TCAGTGAATAATTAACCCACAATCTAGTGGCACAAATTAACATATCTCGGGAATATAACA 2280
2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGG 2340
2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGG 2340
2341 AACGTCACTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTATTAATACATTTCA 2400
2341 AACGTCACTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTATTAATACATTTCA 2400
2401 AGCAATAGCAAGGCTTAACACACACAGTATAGAAGCTCTCGAGGGGTGAATTTTAAACGGC 2460
2401 AGCAATAGCAAGGCTTAACACACACAGTATAGAAGCTCTCGAGGGGTGAATTTTAAACGGC 2460
2461 GTAATGGCAACATGCTCAATCTCAAGAGAGGAGGAGGAACTTAATTTCAAAATTAATA 2520
2461 GTAATGGCAACATGCTCAATCTCAAGAGAGGAGGAGGAACTTAATTTCAAAATTAATA 2520
2521 CCAACAGAGACATGAACACACACACCTTTACCAATTCGGTTTATAGCCAAATACACA 2580
2521 CCAACAGAGACATGAACACACACACCTTTACCAATTCGGTTTATAGCCAAATACACA 2580
2581 GCCACTGGTGGGGCTCTGTTTTTATATATATATGCCAACCATTCCTGGCAGAGGGCT 2640
2581 GCCACTGGTGGGGCTCTGTTTTTATATATATATGCCAACCATTCCTGGCAGAGGGCT 2640
2641 GAGTTAAATAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTTAAATTCOCAT 2700
2641 GAGTTAAATAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTTAAATTCOCAT 2700
2701 GTTCGGCGGATGACGCTTTTAAATCAACAAAGACATTAACCATTAATGCAACCAATTC 2760
2701 GTTCGGCGGATGACGCTTTTAAATCAACAAAGACATTAACCATTAATGCAACCAATTC 2760
2761 AATTCACGCTCAGACAGACCAAGATGATTTTATGACGGGTACGACGCAATGGCCATC 2820
2761 AATTCACGCTCAGACAGACCAAGATGATTTTATGACGGGTACGACGCAATGGCCATC 2820
2821 AATTCACGCTCAGACATATCCATCTGGCGGTATGTCACCCCTGGTGGCAACAACTCA 2880
2821 AATTCACGCTCAGACATATCCATCTGGCGGTATGTCACCCCTGGTGGCAACAACTCA 2880
2881 AGCAGCAGATACGGGAATATTAATCTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2940
2881 AGCAGCAGATACGGGAATATTAATCTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2940
2941 AATAAGCCCTTAATCAGCAAAACATTAAGGATAGAGTTATAAATTTTAAAGCTTGGCTG 3000
2941 AATAAGCCCTTAATCAGCAAAACATTAAGGATAGAGTTATAAATTTTAAAGCTTGGCTG 3000
3001 GTTAATGGAGTTTAAAGTTAACTGGCGAAATGAGATATTAAGGCAATCTCACTAT 3060
3001 GTTAATGGAGTTTAAAGTTAACTGGCGAAATGAGATATTAAGGCAATCTCACTAT 3060
3061 TCAGAAAGCCCACTTTTAAAGGAAGACTAGAGATACCTTAATATCACCGCAATTTT 3120
3061 TCAGAAAGCCCACTTTTAAAGGAAGACTAGAGATACCTTAATATCACCGCAATTTT 3120
3121 ACCAATATGGCACTGCCGAAATTAATATAACACAGGAGTGGTAAACCTTGGCAATGTT 3180
3121 ACCAATATGGCACTGCCGAAATTAATATAACACAGGAGTGGTAAACCTTGGCAATGTT 3180
3181 ACCAATGATGGTGAATTAACATTAACATTAACATTAACATTAACATTAACATTAACAT 3240
3181 ACCAATGATGGTGAATTAACATTAACATTAACATTAACATTAACATTAACATTAACAT 3240
3241 GCGGAGATATATCAACAAAGAGGAGCTTTAAATTTTACAGACAGTAAATATGATGCT 3300
3241 GCGGAGATATATCAACAAAGAGGAGCTTTAAATTTTACAGACAGTAAATATGATGCT 3300
3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAGAGGCAACCTTCAGATTTCTCCGAT 3360

3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAGGCAACCTTCAGATTTCTCCGAT 3360
3361 AAAATTAATATACCAAAACAGATACAAATCAAAAGGCTATTGATGGAGAGGACTCTAGT 3420
3361 AAAATTAATATACCAAAACAGATACAAATCAAAAGGCTATTGATGGAGAGGACTCTAGT 3420
3421 TCAGATGGCAGCAAGTAATGGCCAACTTAATTTAAACCAAGAGATTCAAATTCACAGAA 3480
3421 TCAGATGGCAGCAAGTAATGGCCAACTTAATTTAAACCAAGAGATTCAAATTCACAGAA 3480
3481 GACCTAAGTATTTTCAAGGTTTCAATTAAGCAGAGATTTAGACCCAAAGATTTAGAGATT 3540
3481 GACCTAAGTATTTTCAAGGTTTCAATTAAGCAGAGATTTAGACCCAAAGATTTAGAGATT 3540
3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTGCAGGCGCAAAACAGTAACTTTTAAAC 3600
3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTGCAGGCGCAAAACAGTAACTTTTAAAC 3600
3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGCTCAAAATGTGACACTAAATAGCAAGTG 3660
3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGCTCAAAATGTGACACTAAATAGCAAGTG 3660
3661 AAAAAATCTAGCAGCAATGCGGACGTAAAGCAATAGCAGCAAGATACCGGCTTAACT 3720
3661 AAAAAATCTAGCAGCAATGCGGACGTAAAGCAATAGCAGCAAGATACCGGCTTAACT 3720
3721 ATTACTGCAAAAATGTAGAAGTAACAAAGATATTACTTCTCAAAAACAGTAAATATC 3780
3721 ATTACTGCAAAAATGTAGAAGTAACAAAGATATTACTTCTCAAAAACAGTAAATATC 3780
3781 ACCGCTCGGAAAAGTTACCCACAGCAGAGGCTCGACCATTAACCCACAAATGCGCAA 3840
3781 ACCGCTCGGAAAAGTTACCCACAGCAGAGGCTCGACCATTAACCCACAAATGCGCAA 3840
3841 GCAAGTATTACAAACCAAGAGTATACAGCGGTACGATTTCCCGGTAAACAGGTAAGT 3900
3841 GCAAGTATTACAAACCAAGAGTATACAGCGGTACGATTTCCCGGTAAACAGGTAAGT 3900
3901 GTTAGCGGACTGGTGAATTTAAACCACTAAATCCGGGTCAAAAATTTGAAGGAAATCGG 3960
3901 GTTAGCGGACTGGTGAATTTAAACCACTAAATCCGGGTCAAAAATTTGAAGGAAATCGG 3960
3961 GAGGCTAATTAACCAAGTCAACAGAGTACAAATTTGGGGTACAAATTTCCCGTAAATAC 4020
3961 GAGGCTAATTAACCAAGTCAACAGAGTACAAATTTGGGGTACAAATTTCCCGTAAATAC 4020
4021 AATGTTACGCAACGCTGGCGATTTTAAACAGTTGGGAAATGGCGCAAGATTAATGGCA 4080
4021 AATGTTACGCAACGCTGGCGATTTTAAACAGTTGGGAAATGGCGCAAGATTAATGGCA 4080
4081 GAAGGCTGCAACCTTTAAACGCAACAGGAAATACCTTACTTAAAGCGGTTCTAGC 4140
4081 GAAGGCTGCAACCTTTAAACGCAACAGGAAATACCTTACTTAAAGCGGTTCTAGC 4140
4141 ATCACTTCAACTAAGGTCAGGTAGCTCTGGCTCAGATGATGATGATGATGATGATGATG 4200
4141 ATCACTTCAACTAAGGTCAGGTAGCTCTGGCTCAGATGATGATGATGATGATGATGATG 4200
4201 ATTAATGCTGCTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
4201 ATTAATGCTGCTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
4261 GATATTAAGCAACACAGCGGCACTTGGTTTAAACGCAAAAGATGCTTAAGCTAAATG 4320
4261 GATATTAAGCAACACAGCGGCACTTGGTTTAAACGCAAAAGATGCTTAAGCTAAATG 4320
4321 GATGATCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4380
4321 GATGATCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4380
4381 ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTTAAACACAGTAAATGGGTT 4440
4381 ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTTAAACACAGTAAATGGGTT 4440

```

QY 4441 AATATCATTTTCGAAGATGGTAGAAACACTGTGCGCTTAAGAGGCAAGAAATTTAGAGTG 4500
Db 4441 AATATCATTTTCGAAGATGGTAGAAACACTGTGCGCTTAAGAGGCAAGAAATTTAGAGTG 4500
QY 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAAGAAGTAAATTTGAAGGCAACGCGTCCTT 4560
Db 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAAGAAGTAAATTTGAAGGCAACGCGTCCTT 4560
QY 4561 GAAAGAGTAAAGATTTATCTGTATGAAGAAGAGAAACATTTAGCTAAACTTGGTGTAAAGT 4620
Db 4561 GAAAGAGTAAAGATTTATCTGTATGAAGAAGAGAAACATTTAGCTAAACTTGGTGTAAAGT 4620
QY 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
Db 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
QY 4681 ACCAGCCGTCAAGTCAAGTATTTCTGGAAGGTAAAGGCGTGTCTCTCAAGTGGTAAAT 4740
Db 4681 ACCAGCCGTCAAGTCAAGTATTTCTGGAAGGTAAAGGCGTGTCTCTCAAGTGGTAAAT 4740
QY 4741 GCGCAGCAGTATGTACCAATGTTGCTGACGATGACAGCGCTAGTCAGTAAATTCACAAAG 4800
Db 4741 GCGCAGCAGTATGTACCAATGTTGCTGACGATGACAGCGCTAGTCAGTAAATTCACAAAG 4800
QY 4801 GTAGATTTTCATCTGCAATGAAGTCAATTTTATTTTCTGATATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTTCATCTGCAATGAAGTCAATTTTATTTTCTGATATTTACTGTGGGTTAAA 4860
QY 4861 GTTCAGTACGGGCTTACCCTATCTTGTAAATAATTTACGGAGATACAAATAAGTATTTT 4920
Db 4861 GTTCAGTACGGGCTTACCCTATCTTGTAAATAATTTACGGAGATACAAATAAGTATTTT 4920
QY 4921 AACAGGTTATTATTATG 4937
Db 4921 AACAGGTTATTATTATG 4937

RESULT 2
US-08-302-832-3
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651

```

```

; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-3

```

```

Query Match      100.0%; Score 4937; DB 1; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TAAATATACAGATATAATAAATAAATCAAGATTTTTGTGATGACAAACAACAATTACAA 60
Db 1 TAAATATACAGATATAATAAATAAATCAAGATTTTTGTGATGACAAACAACAATTACAA 60
QY 61 CACCTTTTTCAGCTTATATGCAAAATATTTTAAAAAATAAGTATAAATCCGCCATATA 120
Db 61 CACCTTTTTCAGCTTATATGCAAAATATTTTAAAAAATAAGTATAAATCCGCCATATA 120
QY 121 AATGGTATAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTT 180
Db 121 AATGGTATAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTT 180
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 240
Db 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 240
QY 241 CACATGAAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 241 CACATGAAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
QY 301 GAACGCAATGATGAAGTAAATTTAATTTGTTCACATAACCTTTAGGAGAAATATGAACA 360
Db 301 GAACGCAATGATGAAGTAAATTTAATTTGTTCACATAACCTTTAGGAGAAATATGAACA 360
QY 361 ATATATCGTCTCAAAATTCAGCAAAAGCGCTGAAATCTTTGGTGTCTGTCTGAAATGGCA 420
Db 361 ATATATCGTCTCAAAATTCAGCAAAAGCGCTGAAATCTTTGGTGTCTGTCTGAAATGGCA 420
QY 421 CGGGGTGTGACCAATTCACAGAAAAGCGTTCGGCTATGTACTATCTTTTAGGTGTAAAC 480
Db 421 CGGGGTGTGACCAATTCACAGAAAAGCGTTCGGCTATGTACTATCTTTTAGGTGTAAAC 480
QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTACTATCTTTAGGTGTAAACATCTATTCCA 540
Db 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTACTATCTTTAGGTGTAAACATCTATTCCA 540
QY 541 CAATCTGTTTACGACGCGCTTTACAAGGAATGGATGTAGTACAGGACAGGCAAGCTATG 600
Db 541 CAATCTGTTTACGACGCGCTTTACAAGGAATGGATGTAGTACAGGACAGGCAAGCTATG 600
QY 601 CAAGTAGATGGTAAATAAACCAATTTATCCGCAACAGTGTTCAGCGCTATCAATTTGAAA 660
Db 601 CAAGTAGATGGTAAATAAACCAATTTATCCGCAACAGTGTTCAGCGCTATCAATTTGAAA 660
QY 661 CAATTTAATATCGACCAAAATGAATGGTGGCTTTTACAGAAAACAACAACACTCCGCC 720
Db 661 CAATTTAATATCGACCAAAATGAATGGTGGCTTTTACAGAAAACAACAACACTCCGCC 720
QY 721 GTATTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGATTTTAGATTCTAAC 780
Db 721 GTATTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGATTTTAGATTCTAAC 780
QY 781 GGACAAAGTCTTTTAAATCAACCCCAATGTATCACAATAGGTAAAGACGAATTTATTAC 840
Db 781 GGACAAAGTCTTTTAAATCAACCCCAATGTATCACAATAGGTAAAGACGAATTTATTAC 840

```

QY 841 ACTAATGGCTTTACGGCTTCTACGGTAGACATTTCTAAGCAAAACATCAAGCGCGTAAAT 900
D 841 ACTAATGGCTTTACGGCTTCTACGGTAGACATTTCTAAGCAAAACATCAAGCGCGTAAAT 900
QY 901 TACACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACCGTTTAAAT 960
D 901 TACACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACCGTTTAAAT 960
QY 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGTTGCAAAAGTGAANAACAGAGGTGTG 1020
D 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGTTGCAAAAGTGAANAACAGAGGTGTG 1020
QY 1021 ATTAGCGTAAATGCTGGCAGCATTTCTTACTGCGAGGCAAAATATCAACATCAGCGAT 1080
D 1021 ATTAGCGTAAATGCTGGCAGCATTTCTTACTGCGAGGCAAAATATCAACATCAGCGAT 1080
QY 1081 ATATAAACCACCACTTACTTACAGCATTTGCGCGCCTGAAATGAAGCGGTCAATCTG 1140
D 1081 ATATAAACCACCACTTACTTACAGCATTTGCGCGCCTGAAATGAAGCGGTCAATCTG 1140
QY 1141 GCGGATATTTTGCRAAGCGGTAACTTAATGTCGCTGCCACTATTCGAACCAAA 1200
D 1141 GCGGATATTTTGCRAAGCGGTAACTTAATGTCGCTGCCACTATTCGAACCAAA 1200
QY 1201 GGTAACTTTCTGCTGATCTGAAGCAAGATAAAGCGGCAATATTTGTTCTTCGCG 1260
D 1201 GGTAACTTTCTGCTGATCTGAAGCAAGATAAAGCGGCAATATTTGTTCTTCGCG 1260
QY 1261 AAAGAGGCTGAAGCGGAAATTTGCGGTGTAAATTCGCTCAAAATCAGCAAGCTAAAGC 1320
D 1261 AAAGAGGCTGAAGCGGAAATTTGCGGTGTAAATTCGCTCAAAATCAGCAAGCTAAAGC 1320
QY 1321 GCGAAGCTGTATGATTACAGCGGATAAAGTCAATTTAAAAACAGGTGCGAGTTTCGACCTT 1380
D 1321 GCGAAGCTGTATGATTACAGCGGATAAAGTCAATTTAAAAACAGGTGCGAGTTTCGACCTT 1380
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGCGAGGTAAAAAC 1440
D 1381 TCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGCGAGGTAAAAAC 1440
QY 1441 GCGATCAATTTAGCAAAAGAAACCTTTTAAAGAAAGCTCAACCATCAATGTATCAGC 1500
D 1441 GCGATCAATTTAGCAAAAGAAACCTTTTAAAGAAAGCTCAACCATCAATGTATCAGC 1500
QY 1501 AAAGAAAGCGGACGCGCTATTGTGTTGGCGCATATTGCTTAATTTGACGCGCAATTT 1560
D 1501 AAAGAAAGCGGACGCGCTATTGTGTTGGCGCATATTGCTTAATTTGACGCGCAATTT 1560
QY 1561 AACGCTCAAGGTAGTGTATCGCTAAACCGGTGTTTGTGGAGACATCGGGCAT 1620
D 1561 AACGCTCAAGGTAGTGTATCGCTAAACCGGTGTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTTATCCATTTGACGCAATGCAATTTTAAACAAAGAGTGGTGTGCTAGACCCCTGAT 1680
D 1621 TATTTATCCATTTGACGCAATGCAATTTTAAACAAAGAGTGGTGTGCTAGACCCCTGAT 1680
QY 1681 GATGTAACAATTTGAAGCGGAGACCCCTTCGCAATAATACCGGTATAAATGATGAATTC 1740
D 1681 GATGTAACAATTTGAAGCGGAGACCCCTTCGCAATAATACCGGTATAAATGATGAATTC 1740
QY 1741 CCAACAGGCGGCTGAAGCAAGCGACCTTAAAAAATATAGCAACTCAAAACACGCTA 1800
D 1741 CCAACAGGCGGCTGAAGCAAGCGACCTTAAAAAATATAGCAACTCAAAACACGCTA 1800
QY 1801 ACCAATACACATTTTCAAAATTTCTGAAAGCGCTGACAAATGAATTAACGGCATCA 1860
D 1801 ACCAATACACATTTTCAAAATTTCTGAAAGCGCTGACAAATGAATTAACGGCATCA 1860
QY 1861 AGAAACTTACCGTTTAAAGCTCAATCAGCATCGAAGCACTCCCACTTAATTCCTCAT 1920
D 1861 AGAAACTTACCGTTTAAAGCTCAATCAGCATCGAAGCACTCCCACTTAATTCCTCAT 1920
QY 1921 AGTAAAGGTACGGTGGCGGAGGGTTCAGATTGTGAGATATTACTTCTAAAGCGGA 1980

D 1921 AGTAAAGGTACGGTGGCGGAGGGTTCAGATTGATGAGATATTACTTCTAAAGCGGA 1980
QY 1981 AATTTAACCATTTTATCTGCGGATGGTTGATCTTCATAAAAATATTACGCTTGATCAG 2040
D 1981 AATTTAACCATTTTATCTGCGGATGGTTGATCTTCATAAAAATATTACGCTTGATCAG 2040
QY 2041 GGTTTTTTAAATATACCGCGCTTCCTAGCTTTTGAAGTGGAAATTAACAAAGCACGC 2100
D 2041 GGTTTTTTAAATATACCGCGCTTCCTAGCTTTTGAAGTGGAAATTAACAAAGCACGC 2100
QY 2101 GACGGGCAATGCTAAATTTGTCGCCAGGCGACTGTACCAATTACAGGAGAGGAAAA 2160
D 2101 GACGGGCAATGCTAAATTTGTCGCCAGGCGACTGTACCAATTACAGGAGAGGAAAA 2160
QY 2161 GATTTACGGGTAAACACGATATCTTTAAACGAGGCTAAAGTCTGAATATCATTTCA 2220
D 2161 GATTTACGGGTAAACACGATATCTTTAAACGAGGCTAAAGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATAATTTAACCCACAATCTTAGTGCGCAATTAACATATCTGGGAATAACA 2280
D 2221 TCAGTGAATAATTTAACCCACAATCTTAGTGCGCAATTAACATATCTGGGAATAACA 2280
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGACATGG 2340
D 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGACATGG 2340
QY 2341 AACGTCAGTGTCTTAACTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2400
D 2341 AACGTCAGTGTCTTAACTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACAGTATAGAGCTCTGACAGGCTGAATTTTAAACGCG 2460
D 2401 AGCAATAGCAAGGCTTAAACACAGTATAGAGCTCTGACAGGCTGAATTTTAAACGCG 2460
QY 2461 GTAAATGGCAACATGTCATTCATCTCAAGAGGAGCGCAAGTAAATTTCAATTTAAAA 2520
D 2461 GTAAATGGCAACATGTCATTCATCTCAAGAGGAGCGCAAGTAAATTTCAATTTAAAA 2520
QY 2521 CCAACGAGCAACATGAAACCAAGCAACCTTTTACCAATTCGGTTTACGCAATATCA 2580
D 2521 CCAACGAGCAACATGAAACCAAGCAACCTTTTACCAATTCGGTTTACGCAATATCA 2580
QY 2581 GCCACTGGTGGGGCTCTGTTTTTGTATATATATGCAACCATTTCTGCGAGAGGGCT 2640
D 2581 GCCACTGGTGGGGCTCTGTTTTTGTATATATATGCAACCATTTCTGCGAGAGGGCT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGGCTAAATTTTACCTTTAAATTCCTCAT 2700
D 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGGCTAAATTTTACCTTTAAATTCCTCAT 2700
QY 2701 GTTTCGGGCGATCAGCTTTTAAAAATCAACAAAGACTTAAACCAATAAATGCAACCAATTC 2760
D 2701 GTTTCGGGCGATCAGCTTTTAAAAATCAACAAAGACTTAAACCAATAAATGCAACCAATTC 2760
QY 2761 AATTTACGCTCAGACAGACGAAGATGATTTTATGACGGGTACGACCAATGCGCATC 2820
D 2761 AATTTACGCTCAGACAGACGAAGATGATTTTATGACGGGTACGACCAATGCGCATC 2820
QY 2821 AATTTCAACCTACACATATCCATTCGCGGCTTAATGTCACCTTTGTCGACAAACTCA 2880
D 2821 AATTTCAACCTACACATATCCATTCGCGGCTTAATGTCACCTTTGTCGACAAACTCA 2880
QY 2881 AGCAGCAGCTTACGGGGAATATTACTATCGAGAAAGCAAGCAATGTTAGCTAGAGCC 2940
D 2881 AGCAGCAGCTTACGGGGAATATTACTATCGAGAAAGCAAGCAATGTTAGCTAGAGCC 2940
QY 2941 AATAACGCCCTTAATCAGCAAAACATAAGGATAGAGTTATAAATCTTGGCAGCTTCCCTC 3000
D 2941 AATAACGCCCTTAATCAGCAAAACATAAGGATAGAGTTATAAATCTTGGCAGCTTCCCTC 3000
QY 3001 GTTAAATGGAGCTTTTAAAGTTTAACTGGGGAATGCAGATTTTAAAGGCAATCTCCTATT 3060
D 3001 GTTAAATGGAGCTTTTAAAGTTTAACTGGGGAATGCAGATTTTAAAGGCAATCTCCTATT 3060

Db 3001 GTTAATGGAGTTTAAAGTTTAACTGGCGGAAAATGCAGATATTAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGCGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGCGCAATTTT 3120
QY 3121 ACCAATATGGCACTGCGCAATTAATTAACACAAAGGAGTGGTAAACCTTGGCAATGTT 3180
Db 3121 ACCAATATGGCACTGCGCAATTAATTAACACAAAGGAGTGGTAAACCTTGGCAATGTT 3180
QY 3181 ACCAATGATGGTGAATTAACCAATTTACCACCTACGCTAAACCAAGGAGTGGTAAACCTTGGCAATGTT 3240
Db 3181 ACCAATGATGGTGAATTAACCAATTTACCACCTACGCTAAACCAAGGAGTGGTAAACCTTGGCAATGTT 3240
QY 3241 GCGCGAGATATATCAACAAAAGGAGGCTTAAATATTACAGACAGTAAATATGATGCT 3300
Db 3241 GCGCGAGATATATCAACAAAAGGAGGCTTAAATATTACAGACAGTAAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGGCAATATCTGCAAAAAGGAGGCACTACGATTTCTTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGGCAATATCTGCAAAAAGGAGGCACTACGATTTCTTCCGAT 3360
QY 3361 AAAATTAATATCACCACACAGATTAACAAATCAAAAAGGAGTATTGATGGAGAGCTCTAGT 3420
Db 3361 AAAATTAATATCACCACACAGATTAACAAATCAAAAAGGAGTATTGATGGAGAGCTCTAGT 3420
QY 3421 TCAGATGCGACAAGTAATGCGCAACCTTAACCTATTAAACCAAGGAGTATTGATGGAGAGCTCTAGT 3480
Db 3421 TCAGATGCGACAAGTAATGCGCAACCTTAACCTATTAAACCAAGGAGTATTGATGGAGAGCTCTAGT 3480
QY 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTAACGCCAAAGAGTGGTAGAGTTTA 3540
Db 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTAACGCCAAAGAGTGGTAGAGTTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAAAGCAATAGGCAACAGTAACTTAACT 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAAAGCAATAGGCAACAGTAACTTAACT 3600
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCACAATGTGACACTAAATAGCAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCACAATGTGACACTAAATAGCAAGTG 3660
QY 3661 AAACATCTAGCAGCAATGGCGGAGCTGAAAGCAATAGGCAACAGTAACTTAACT 3720
Db 3661 AAACATCTAGCAGCAATGGCGGAGCTGAAAGCAATAGGCAACAGTAACTTAACT 3720
QY 3721 ATTACTGCAAAAATGTAGAGCTTAACCAAGATATTACTTCTCTCAAAACAGTAAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAGCTTAACCAAGATATTACTTCTCTCAAAACAGTAAATATC 3780
QY 3781 ACCGGTGGGAAAAGTTACCAACAGCAGGCTCGACCAATTAAGCCAAACAAATGGCAAA 3840
Db 3781 ACCGGTGGGAAAAGTTACCAACAGCAGGCTCGACCAATTAAGCCAAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACCAACCAACAGCTGATATCAGCGGTACGATTTCCGGTAAACAGGTAAGT 3900
Db 3841 GCAAGTATTACCAACCAACAGCTGATATCAGCGGTACGATTTCCGGTAAACAGGTAAGT 3900
QY 3901 GTTAGCGGAGCTGGTGAATTAACCACTAAATCCGGCTCAAAAATTAAGCGAATCCGGT 3960
Db 3901 GTTAGCGGAGCTGGTGAATTAACCACTAAATCCGGCTCAAAAATTAAGCGAATCCGGT 3960
QY 3961 GAGGCTAATGTAACAAGTGCACAGGTAACATTTGGCGGTACAAATTTCCGGTAATACGGTA 4020
Db 3961 GAGGCTAATGTAACAAGTGCACAGGTAACATTTGGCGGTACAAATTTCCGGTAATACGGTA 4020
QY 4021 AATGTTACGCAACCGCTGGCGATTAAACAGTTGGGAATGGCGAGGAGTAAATGCGACA 4080
Db 4021 AATGTTACGCAACCGCTGGCGATTAAACAGTTGGGAATGGCGAGGAGTAAATGCGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGAGTAACTTGAATGACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGAGTAACTTGAATGACTACTGAAGCGGTTCTAGC 4140

QY 4141 ATCACTTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCCGCAAGC 4200
Db 4141 ATCACTTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCCGCAAGC 4200
QY 4201 ATTAACTGCTGAATTTGTCAGATTAATTAACAGGAGTGGTAAACCTTGGCAATGTT 4260
Db 4201 ATTAACTGCTGAATTTGTCAGATTAATTAACAGGAGTGGTAAACCTTGGCAATGTT 4260
QY 4261 GATATTAAAGCAACCAAGCGGCACTTGGTTTAAACGCAAAAGATGCTAAGCTAAATGTT 4320
Db 4261 GATATTAAAGCAACCAAGCGGCACTTGGTTTAAACGCAAAAGATGCTAAGCTAAATGTT 4320
QY 4321 GATGCAATCAGGTGATAGTACAGAGTGAATGAGTGAAGTGAAGTGAAGTGAAGT 4380
Db 4321 GATGCAATCAGGTGATAGTACAGAGTGAATGAGTGAAGTGAAGTGAAGTGAAGT 4380
QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTAAACACAGTAAATGGGTTA 4440
Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTAAACACAGTAAATGGGTTA 4440
QY 4441 AATATCATTTGCAAGAGTGTAGAACACACTGTGCGCTTAAGAGGCAAGGAAATTTGAGTG 4500
Db 4441 AATATCATTTGCAAGAGTGTAGAACACACTGTGCGCTTAAGAGGCAAGGAAATTTGAGTG 4500
QY 4501 AATATATATCCAGCGGAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGT 4560
Db 4501 AATATATATCCAGCGGAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGT 4560
QY 4561 GAAAAAGTAAAGAGTTTATCTGATGAGCAAGGAGTAAATGAGGCAAGGAAATTTGAGTG 4620
Db 4561 GAAAAAGTAAAGAGTTTATCTGATGAGCAAGGAGTAAATGAGGCAAGGAAATTTGAGTG 4620
QY 4621 GCTGTACGTTTGTGGAGCAATTAATACAATTAACAGTCAATACACAAATGAAATTTACA 4680
Db 4621 GCTGTACGTTTGTGGAGCAATTAATACAATTAACAGTCAATACACAAATGAAATTTACA 4680
QY 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAGGTAAGGCGGTTTCTCAAGTGAAT 4740
Db 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAGGTAAGGCGGTTTCTCAAGTGAAT 4740
QY 4741 GCGGCAAGTATGATCAATTTGCTGACGATGAGCAGCGGCTAGTCAATTTGACAAG 4800
Db 4741 GCGGCAAGTATGATCAATTTGCTGACGATGAGCAGCGGCTAGTCAATTTGACAAG 4800
QY 4801 GTAGATTTTCACTCTGCAATGAAGTCAATTTTCTGATTTTCTGATTTTCTGATTTTCT 4860
Db 4801 GTAGATTTTCACTCTGCAATGAAGTCAATTTTCTGATTTTCTGATTTTCTGATTTTCT 4860
QY 4861 GTTCTAGTACGGGCTTTACCCATCTTTGTAAGGAGTAAATGAGTAAATTTT 4920
Db 4861 GTTCTAGTACGGGCTTTACCCATCTTTGTAAGGAGTAAATGAGTAAATTTT 4920
QY 4921 AACAGGTTATTTATG 4937
Db 4921 AACAGGTTATTTATG 4937

RESULT 3

US-08-530-198-3
; Sequence 3, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington

Db 1681 GATGTACAAATTGAAGCGGAGAGACCCCTTCGCAATTAATACCGGTATAAATGATGAATTC 1740
QY 1741 CCAACAGCCACCGGTGAAGCAGCGCCCTAAACAAAATAGCGAACTCAACAAACAGCTA 1800
Db 1741 CCAACAGCCACCGGTGAAGCAGCGCCCTAAACAAAATAGCGAACTCAACAAACAGCTA 1800
QY 1801 ACCAATCAAACTATTTCAAAATATCTGAAAAGCGCTGGACAATGAATATACGGCATCA 1860
Db 1801 ACCAATCAAACTATTTCAAAATATCTGAAAAGCGCTGGACAATGAATATACGGCATCA 1860
QY 1861 AGAAAATTAACGTTAATAGCTCAATCAACATCGGAACACCTCCCACTTAATTCGCAT 1920
Db 1861 AGAAAATTAACGTTAATAGCTCAATCAACATCGGAACACCTCCCACTTAATTCGCAT 1920
QY 1921 AGTAAAGCTCAGCGTGGCGGAGCGGTTTCAGATTGATGGAGATATTACTTCTAAAGGGGA 1980
Db 1921 AGTAAAGCTCAGCGTGGCGGAGCGGTTTCAGATTGATGGAGATATTACTTCTAAAGGGGA 1980
QY 1981 AATTTAACCAATTTATCTGGCGGATGGGTTGATGTTTCATATAAATATTACGCTTGATCAG 2040
Db 1981 AATTTAACCAATTTATCTGGCGGATGGGTTGATGTTTCATATAAATATTACGCTTGATCAG 2040
QY 2041 GGTTTTTTAAATATTACCGCGCTTCGCTAGCTTTTGAAGGTGGAATAACAAAGCAGC 2100
Db 2041 GGTTTTTTAAATATTACCGCGCTTCGCTAGCTTTTGAAGGTGGAATAACAAAGCAGC 2100
QY 2101 GACGCGCAAAATGCTFAAAATTTGCGCCAGGCACTGTAAACCAATTAACAGGAGGAAA 2160
Db 2101 GACGCGCAAAATGCTFAAAATTTGCGCCAGGCACTGTAAACCAATTAACAGGAGGAAA 2160
QY 2161 GATTTACAGGCTTAACAGCTATCTTTAAACGGAAGCGGTAAAGCTGTAATATCATTTCA 2220
Db 2161 GATTTACAGGCTTAACAGCTATCTTTAAACGGAAGCGGTAAAGCTGTAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTTAACCCAAATCTTAGTGCCACAATTAACATATCTGGGAATATAACA 2280
Db 2221 TCAGTGAATTAATTTAACCCAAATCTTAGTGCCACAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAAACTACGAGAAAGAACCTCGTATTGGCAAAACAGCCATGATTCGCACTGG 2340
Db 2281 ATTAACCAAACTACGAGAAAGAACCTCGTATTGGCAAAACAGCCATGATTCGCACTGG 2340
QY 2341 AAGTCAGTGCTCTTAATCTAGAGACAGCGCCAAATTTTACCTTTTAAATATCATTTCA 2400
Db 2341 AAGTCAGTGCTCTTAATCTAGAGACAGCGCCAAATTTTACCTTTTAAATATCATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGCTCTGCGGGGTGAATTTACGGC 2460
Db 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGCTCTGCGGGGTGAATTTACGGC 2460
QY 2461 GTAATGGCAACATGTCATTTCAATCTCAAGAGGAGCGAAAGTTTCAAAATTAATA 2520
Db 2461 GTAATGGCAACATGTCATTTCAATCTCAAGAGGAGCGAAAGTTTCAAAATTAATA 2520
QY 2521 CCAACGAGAACATGAACACAGCAAGCAACCTTTACCAATTCGGTTTTTAGCCAAATATCA 2580
Db 2521 CCAACGAGAACATGAACACAGCAAGCAACCTTTACCAATTCGGTTTTTAGCCAAATATCA 2580
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGGATATATATGCCAACCTTTCTGCGAGGGGT 2640
Db 2581 GCCACTGGTGGGGCTCTGTTTTTTTGGATATATATGCCAACCTTTCTGCGAGGGGT 2640
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTTAAGCGGCTTAATTTTACCTTAATTCCTCAT 2700
Db 2641 GAGTTAAAAATGAGTGAATTAATATCTTAAGCGGCTTAATTTTACCTTAATTCCTCAT 2700
QY 2701 GTTCGGCGCATGACGCTTTTAAATACAAAAGACTTAACCATATAATGCAACCAATTTCA 2760
Db 2701 GTTCGGCGCATGACGCTTTTAAATACAAAAGACTTAACCATATAATGCAACCAATTTCA 2760
QY 2761 AATTTACGCTCAGACAGCAAGAGATGATTTTTTATGCGGTACGCAACCAATGCAATC 2820
Db 2761 AATTTACGCTCAGACAGCAAGAGATGATTTTTTATGCGGTACGCAACCAATGCAATC 2820

QY 2821 AATTCACCTACACATATTCATTTCTGGCGGTAAATGTACCCCTTGGTGAGAAAACTCA 2880
Db 2821 AATTCACCTACACATATTCATTTCTGGCGGTAAATGTACCCCTTGGTGAGAAAACTCA 2880
QY 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAAGCAGCAATGTACGCTAGAGCC 2940
Db 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAAGCAGCAATGTACGCTAGAGCC 2940
QY 2941 AATAAGCCCTTAATCAGCAAAACATAAGGATAGAGTTATAAAGCTTGGCAGCTTGC 3000
Db 2941 AATAAGCCCTTAATCAGCAAAACATAAGGATAGAGTTATAAAGCTTGGCAGCTTGC 3000
QY 3001 GTTAATGGGAGTTTAAGCTTTAACTGGGAAATGAGATATTAAAGGCAATCTCACATTT 3060
Db 3001 GTTAATGGGAGTTTAAGCTTTAACTGGGAAATGAGATATTAAAGGCAATCTCACATTT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCACCGGCAATTTT 3120
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCACCGGCAATTTT 3120
QY 3121 ACCAATATGGCAGCTCCGCAATTAATATAACAAAGAGTGGTAAAGCTTGGCAATGTT 3180
Db 3121 ACCAATATGGCAGCTCCGCAATTAATATAACAAAGAGTGGTAAAGCTTGGCAATGTT 3180
QY 3181 ACCAATGATGGTGAATTTAAACATTTACCCTAACCTAACGCAACCAAGAGCATCATC 3240
Db 3181 ACCAATGATGGTGAATTTAAACATTTACCCTAACCTAACGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAAAGAGCTTAAATATAACAAAGAGTGGTAAAGCTTGGCAATGTT 3300
Db 3241 GCGGAGATATAATCAACAAAAAGAGCTTAAATATAACAAAGAGTGGTAAAGCTTGGCAATGTT 3300
QY 3301 GAAATCCCAATTTGCGGCAATATCTCGCAAAAGAGGAAAGCTTACGATTTCTTCGAT 3360
Db 3301 GAAATCCCAATTTGCGGCAATATCTCGCAAAAGAGGAAAGCTTACGATTTCTTCGAT 3360
QY 3361 AANAATATATCAACCAAGATTAACAAATCAAAAGAGTATTGATGGAGAGCTCTAGT 3420
Db 3361 AANAATATATCAACCAAGATTAACAAATCAAAAGAGTATTGATGGAGAGCTCTAGT 3420
QY 3421 TCAGATGGCAGCAAGTAAATGCCAACCTTAATTTAAACCAAAAGAAATTCAGATTTAC 3480
Db 3421 TCAGATGGCAGCAAGTAAATGCCAACCTTAATTTAAACCAAAAGAAATTCAGATTTAC 3480
QY 3481 GACCTAAGTATTTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATTTAGAGATT 3540
Db 3481 GACCTAAGTATTTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATTTAGAGATT 3540
QY 3541 ACTATTGCCAACAGTAAATGACGCTTAACAGCGGTGCCGAAGCCAAACAGTAACTTTAAC 3600
Db 3541 ACTATTGCCAACAGTAAATGACGCTTAACAGCGGTGCCGAAGCCAAACAGTAACTTTAAC 3600
QY 3601 AATGTTAAGATTCACAAATCTCTGCTGACGGTCAACAATGTGACACTAAATAGCAAGTG 3660
Db 3601 AATGTTAAGATTCACAAATCTCTGCTGACGGTCAACAATGTGACACTAAATAGCAAGTG 3660
QY 3661 AAACATCTAGCAGCAATGGGAGCTGAAAGCAATAGCAGCAACGATACCGCTTTAAT 3720
Db 3661 AAACATCTAGCAGCAATGGGAGCTGAAAGCAATAGCAGCAACGATACCGCTTTAAT 3720
QY 3721 ATTACTGCAAAAAATGTAGAGTAAACAAAGATTTACTTCTCAAAACAGTAAATATC 3780
Db 3721 ATTACTGCAAAAAATGTAGAGTAAACAAAGATTTACTTCTCAAAACAGTAAATATC 3780
QY 3781 ACCGCTCGGAAAGGTTTACCACAGAGGCTGACCATTTACGCAACCAATGGCAAA 3840
Db 3781 ACCGCTCGGAAAGGTTTACCACAGAGGCTGACCATTTACGCAACCAATGGCAAA 3840
QY 3841 GCAAGTATTACACCAAAAGGTTATACGCGGTACGATTTCCGGTAAACAGGTAAGT 3900
Db 3841 GCAAGTATTACACCAAAAGGTTATACGCGGTACGATTTCCGGTAAACAGGTAAGT 3900

QY 3901 GTTAGCGGACCTGGTGATTTAAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATTCGGGT 3960
Db 3901 GTTAGCGGACCTGGTGATTTAAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATTCGGGT 3960
QY 3961 GAGGCTTAATGTAAACAGTCAACAGGTACAATTTGGCGGTACAAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGGCTTAATGTAAACAGTCAACAGGTACAATTTGGCGGTACAAATTTCCGGTAAATACGGTA 4020
QY 4021 AATGTTTACGCGAAACGCTGGCGATTTAAACAGTTTGGGAATGGCGCAAAATTAATCGACA 4080
Db 4021 AATGTTTACGCGAAACGCTGGCGATTTAAACAGTTTGGGAATGGCGCAAAATTAATCGACA 4080
QY 4081 GAAGGAGCTGCAACCTTTAACCGCAACAGGGAATACCTTCACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTTAACCGCAACAGGGAATACCTTCACTACTGAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTAGCATCGCAGAAGC 4200
Db 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTAGCATCGCAGAAGC 4200
QY 4201 ATTAATGCTGCTAATGTGACATTAATTAATACAGGCACTTAACACCGGTGGCAGGCTCG 4260
Db 4201 ATTAATGCTGCTAATGTGACATTAATTAATACAGGCACTTAACACCGGTGGCAGGCTCG 4260
QY 4261 GATATTAAGCAACAGCGGCACCTTGGTGTATTAAACGCAAAAGATGCTAAGCTAAATGTT 4320
Db 4261 GATATTAAGCAACAGCGGCACCTTGGTGTATTAAACGCAAAAGATGCTAAGCTAAATGTT 4320
QY 4321 GATGATCAGGTGATAGTACAGAGTGAATGCACTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATGATCAGGTGATAGTACAGAGTGAATGCACTCAACGCAAGCGGCTCTGGTAGTGTG 4380
QY 4381 ACTCGGCAACCTCAAGCACTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440
Db 4381 ACTCGGCAACCTCAAGCACTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440
QY 4441 AATATCATTTTCGAAGATGTTAGAACACCTGTGCGCTTAAAGAGCGCAAGAAATTTGAGGTG 4500
Db 4441 AATATCATTTTCGAAGATGTTAGAACACCTGTGCGCTTAAAGAGCGCAAGAAATTTGAGGTG 4500
QY 4501 AATATATCCAGCGGTGAGCAAGTGTAGCAAGTAAATGAAGCAAGCAAGCGGCTCTT 4560
Db 4501 AATATATCCAGCGGTGAGCAAGTGTAGCAAGTAAATGAAGCAAGCAAGCGGCTCTT 4560
QY 4561 GAAAGAGTAAAGATTTATCTGTGAGCAAGAAAGAAACATTAAGTAACTTGGTGAAGT 4620
Db 4561 GAAAGAGTAAAGATTTATCTGTGAGCAAGAAAGAAACATTAAGTAACTTGGTGAAGT 4620
QY 4621 GCTGTACGTTTGTGAGCCAAATTAATACAAATTAAGTAACTTGAAGCAAGTAAATTTACA 4680
Db 4621 GCTGTACGTTTGTGAGCCAAATTAATACAAATTAAGTAACTTGAAGCAAGTAAATTTACA 4680
QY 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGGTGAAGCGGTGTTTCTCAAGTGGTAAT 4740
Db 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGGTGAAGCGGTGTTTCTCAAGTGGTAAT 4740
QY 4741 GCGCACAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTAAATTTGACAAG 4800
Db 4741 GCGCACAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTAAATTTGACAAG 4800
QY 4801 GTAGATTTTCACTTCAATGAAGTCATTTTATTTTCTGATTTATTTCTGTTGGTGTAAA 4860
Db 4801 GTAGATTTTCACTTCAATGAAGTCATTTTATTTTCTGATTTATTTCTGTTGGTGTAAA 4860
QY 4861 GTTCACTAGCGGCTTTACCCATCTGTAAATAATTCGAGAAATACAAATTAAGTATTTT 4920
Db 4861 GTTCACTAGCGGCTTTACCCATCTGTGTAAATAATTCGAGAAATACAAATTAAGTATTTT 4920
QY 4921 AACAGGTTATTAATG 4937
Db 4921 AACAGGTTATTAATG 4937

RESULT 4
US-08-469-880-3
; Sequence 3, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-3
Query Match 100.0%; Score 4937; DB 2; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAATATACAGATAATAATAATAATCAAGATTTTGTGATGACAAACAAACAAATTCACAA 60
Db 1 TAAATATACAGATAATAATAATAATCAAGATTTTGTGATGACAAACAAACAAATTCACAA 60
QY 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120
Db 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120
QY 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
Db 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
QY 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
Db 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
QY 241 CACATGAATGATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db 241 CACATGAATGATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300

Db 4621 GCTGTACGTTTGTGAGCCAAATATACAAATACAGTCAATACACAAAATGAATTTACA 4680
 QY 4681 ACCGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAT 4740
 Db 4681 ACCGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAT 4740
 QY 4741 GCGCAGCAGTATGACCAATGCTGACGATGGACGCCGTAGTCAGTAATGACAAAG 4800
 Db 4741 GCGCAGCAGTATGACCAATGCTGACGATGGACGCCGTAGTCAGTAATGACAAAG 4800
 QY 4801 SPAGATTTCATCCTCAATGAAGTCATTTTATTTTCGTAATTTACTGTGTGGGTAAA 4860
 Db 4801 SPAGATTTCATCCTCAATGAAGTCATTTTATTTTCGTAATTTACTGTGTGGGTAAA 4860
 QY 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAAATAAGTATTTT 4920
 Db 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAAATAAGTATTTT 4920
 QY 4921 AACAGTTATTATTATG 4937
 Db 4921 AACAGTTATTATTATG 4937

RESULT 5

US-08-728-470-3
 ; Sequence 3, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,470
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-633
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4937 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-728-470-3

Query Match 100.0%; Score 4937; DB 2; Length 4937;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TAAATATACAAATATAATAAATAAATCAAGATTTTGTGATGACAAACAACAATTTACAA 60
 Db 1 TAAATATACAAATATAATAAATAAATCAAGATTTTGTGATGACAAACAACAATTTACAA 60
 QY 61 CACCTTTTGTGAGTCTATATGCAAAATTTTAAAAAATAGTATATAATCCGCCATATAA 120
 Db 61 CACCTTTTGTGAGTCTATATGCAAAATTTTAAAAAATAGTATATAATCCGCCATATAA 120
 QY 121 AATGTAATATCTTTCACTTTTCACTTTTAACTTTTCACTTTTCACTTTTCACTTTTCA 180
 Db 121 AATGTAATATCTTTCACTTTTCACTTTTAACTTTTCACTTTTCACTTTTCACTTTTCA 180
 QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
 Db 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
 QY 241 CACATGAAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 Db 241 CACATGAAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 QY 301 CAAACGAAATGATAAAGTAATTTAATTTGTTCAACTAACCTTAGGAGAAAATATGAACA 360
 Db 301 CAAACGAAATGATAAAGTAATTTAATTTGTTCAACTAACCTTAGGAGAAAATATGAACA 360
 QY 361 ATATATCTGTCACAAATTCAGCAAAACGCTGAATGCTTTGGTGTCTGTCGAATTTGCA 420
 Db 361 ATATATCTGTCACAAATTCAGCAAAACGCTGAATGCTTTGGTGTCTGTCGAATTTGCA 420
 QY 421 CGGGTGTGACCAATTCACAGAAAAGGCTTCCGCTATGTTAGTGTACATCTTTTAGGTG 480
 Db 421 CGGGTGTGACCAATTCACAGAAAAGGCTTCCGCTATGTTAGTGTACATCTTTTAGGTG 480
 QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACATCTTTTAGTGTACATCTTTATTTCA 540
 Db 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACATCTTTTAGTGTACATCTTTATTTCA 540
 QY 541 CAATCTGTTTGTAGCAAGCGGCTTCAAGAAATGGATGTAGTACACGGCACAGCCACTATG 600
 Db 541 CAATCTGTTTGTAGCAAGCGGCTTCAAGAAATGGATGTAGTACACGGCACAGCCACTATG 600
 QY 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTTAATTTGAAA 660
 Db 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTTAATTTGAAA 660
 QY 661 CAATTTAACATCGACCAAAATGAAATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 720
 Db 661 CAATTTAACATCGACCAAAATGAAATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 720
 QY 721 GTATTCAACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTACATTTTAAAC 780
 Db 721 GTATTCAACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTACATTTTAAAC 780
 QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATACAAATAGTAAAGAGCGCAATTTAATTA 840
 Db 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATACAAATAGTAAAGAGCGCAATTTAATTA 840
 QY 841 ACTAATGGCTTTACGGCTTTTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 900
 Db 841 ACTAATGGCTTTACGGCTTTTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 900
 QY 901 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATTCACGGTTTAA 960
 Db 901 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATTCACGGTTTAA 960
 QY 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTTGTCGCAAGTGAAGGAGGAGGTTG 1020
 Db 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTTGTCGCAAGTGAAGGAGGAGGTTG 1020

1021	ATTAGCGTAAATGCTGGCAGCAATTTCTTTTACTCGCAGGCGAAAAATCACCATCAGCGAT	1080
1021	ATTAGCGTAAATGCTGGCAGCAATTTCTTTTACTCGCAGGCGAAAAATCACCATCAGCGAT	1080
1081	ATAATAAACCCAAACCAATTACTTACAGCATTTGCCGCGCCTGAAATGAAGCGTCAATCTG	1140
1081	ATAATAAACCCAAACCAATTACTTACAGCATTTGCCGCGCCTGAAATGAAGCGTCAATCTG	1140
1141	GGCGATATTTTTGCCAAAGGCGGTAAACATTAATCTCCGTCGCCCACTATTTCGAAACCAA	1200
1141	GGCGATATTTTTGCCAAAGGCGGTAAACATTAATCTCCGTCGCCCACTATTTCGAAACCAA	1200
1201	GGTAAACTTTCTGCTGATTCTGTAAGCAAAAGATAAAAAGCGGCAATATTGTTCCTTCGCGC	1260
1201	GGTAAACTTTCTGCTGATTCTGTAAGCAAAAGATAAAAAGCGGCAATATTGTTCCTTCGCGC	1260
1261	AAAGAGGCTGAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC	1320
1261	AAAGAGGCTGAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC	1320
1321	GGCAAGCTGATGATTTACAGCGGATAAAGTCACATTAACACAGTGCAGTTATCGACCTT	1380
1321	GGCAAGCTGATGATTTACAGCGGATAAAGTCACATTAACACAGTGCAGTTATCGACCTT	1380
1381	TCAGGTTAAAGAGGGGGAACACTTTACCTTGGCGCTGACGAGCGCGCGAGGTAAAAAC	1440
1381	TCAGGTTAAAGAGGGGGAACACTTTACCTTGGCGGTACGAGCGCGCGAGGTAAAAAC	1440
1441	GGCATTTCAATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGC	1500
1441	GGCATTTCAATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGC	1500
1501	AAAGAAAAAGGCGAGCGCCTATTGTGTGGCGGATATGCGTTAAATGACGGCAATATT	1560
1501	AAAGAAAAAGGCGAGCGCCTATTGTGTGGCGGATATGCGTTAAATGACGGCAATATT	1560
1561	AACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGTTTTGTGGAGACATCGGGGCAT	1620
1561	AACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGTTTTGTGGAGACATCGGGGCAT	1620
1621	TATTTATCCATTGACAGCAATGCAATTTGTATAAACAAAAGAGTGGTCTAGACCCCTGAT	1680
1621	TATTTATCCATTGACAGCAATGCAATTTGTATAAACAAAAGAGTGGTCTAGACCCCTGAT	1680
1681	GATGTTAAACAATTTGAAGCGGAAAGACCCCTTCGCAATATAACGGGTATAAATGATGAATTC	1740
1681	GATGTTAAACAATTTGAAGCGGAAAGACCCCTTCGCAATATAACGGGTATAAATGATGAATTC	1740
1741	CCAAAGGCAACCGGTGAAGCAAGCGACCCCTTAAAAAAAATAGCGAACTCAAAACCAACGCTA	1800
1741	CCAAAGGCAACCGGTGAAGCAAGCGACCCCTTAAAAAAAATAGCGAACTCAAAACCAACGCTA	1800
1801	ACCAATACAACATATTTCAAAATTTATCTGAAAAAGCGCTGGACAATGAATATAACGGCATCA	1860
1801	ACCAATACAACATATTTCAAAATTTATCTGAAAAAGCGCTGGACAATGAATATAACGGCATCA	1860
1861	AGAAAACCTTACGGTTATAGCTCAATCAATCGGAAGCAACTCCACCTTAATTTCTCCAT	1920
1861	AGAAAACCTTACGGTTATAGCTCAATCAATCGGAAGCAACTCCACCTTAATTTCTCCAT	1920
1921	AGTAAAGTCAAGCTGGCGAGCGGTTACAGATTGATGGAGATATTACTTCTAAAGCGGA	1980
1921	AGTAAAGTCAAGCTGGCGAGCGGTTACAGATTGATGGAGATATTACTTCTAAAGCGGA	1980
1981	AATTTAACCAATTTATCTGGCGGATGGTTGATGTTTCATAAAAATATTACGCTTGATCAG	2040
1981	AATTTAACCAATTTATCTGGCGGATGGTTGATGTTTCATAAAAATATTACGCTTGATCAG	2040
2041	GGTTTTTTTAAATATTACCGCGCTTCGCTAGCTTTTGAAGGTGGAATAACAAGCACGC	2100
2041	GGTTTTTTTAAATATTACCGCGCTTCGCTAGCTTTTGAAGGTGGAATAACAAGCACGC	2100
2101	GACGCGCGCAATTGTCTAAATTTGTCGCCCGAGGGCACTTAACCATTTACGAAGAGGAAAA	2160

2101	g		GAGCGGCAATGCTAAATGTGCGCCAGGCGCACTGTAACCATTTACAGAGAGGGAAAA	2160
2161	y		GAITTCAGGGCTAACACGATATCTTTAAACGGACGGGTAAAGGTCCTGAATATCATTTCA	2220
2161	b		GAITTCAGGGCTAACACGATATCTTTAAACGGACGGGTAAAGGTCCTGAATATCATTTCA	2220
2221	y		TCAGTGAATAATTTAAACCCACAATCTTAGTGGCACAAATTAACATATCTCGGAATATAACA	2280
2221	b		TCAGTGAATAATTTAAACCCACAATCTTAGTGGCACAAATTAACATATCTCGGAATATAACA	2280
2281	y		ATTAAACAAACTACGAGAAAGAACACCTTCGTATTGGGCAAAACAGCCATGATTCGGACTGG	2340
2281	b		ATTAAACAAACTACGAGAAAGAACACCTTCGTATTGGGCAAAACAGCCATGATTCGGACTGG	2340
2341	y		AACGTCAGTGCCTTTAATCTAGAGACAGGCGCAAAATTTTACCTTTTAAATACATTTCA	2400
2341	b		AACGTCAGTGCCTTTAATCTAGAGACAGGCGCAAAATTTTACCTTTTAAATACATTTCA	2400
2401	y		AGCAATTAGCAAAAGGCTTAAACAACACAGTATAGAAGCTCTGCAAGGGTGAATTTTAACGGC	2460
2401	b		AGCAATTAGCAAAAGGCTTAAACAACACAGTATAGAAGCTCTGCAAGGGTGAATTTTAACGGC	2460
2461	y		GTAATTTGSCAACATGTCATTTCAATCTCAAGAAAGGAGCGAAAGTTAATTTCAAAATTTAAA	2520
2461	b		GTAATTTGSCAACATGTCATTTCAATCTCAAGAAAGGAGCGAAAGTTAATTTCAAAATTTAAA	2520
2521	y		CGAAACGAGAACATGAACACAGCAAAACCTTTTACCAATTCGGTTTTTAGCCAAATATCACA	2580
2521	b		CGAAACGAGAACATGAACACAGCAAAACCTTTTACCAATTCGGTTTTTAGCCAAATATCACA	2580
2581	y		GCCACTGGTGGGGGCTCTGTTTTTTTGATATATATGCCAACCACTTCGCGCAGAGGGGCT	2640
2581	b		GCCACTGGTGGGGGCTCTGTTTTTTTGATATATATGCCAACCACTTCGCGCAGAGGGGCT	2640
2641	y		GAGTTAAAAATGAGTGAAATTAATATCTCTACGGGGCTAATTTTACCTTAATTTCCCAT	2700
2641	b		GAGTTAAAAATGAGTGAAATTAATATCTCTACGGGGCTAATTTTACCTTAATTTCCCAT	2700
2701	y		GTTTCGGCGCGATGACGCTTTTAAAAATCAACAAAGACTTAACCATTAATTCGAACCAATTTCA	2760
2701	b		GTTTCGGCGCGATGACGCTTTTAAAAATCAACAAAGACTTAACCATTAATTCGAACCAATTTCA	2760
2761	y		AAITTCAGCCTCAGACAGAGCAAGATGATTTTATGACGGGTACGACGCAATGCCATC	2820
2761	b		AAITTCAGCCTCAGACAGAGCAAGATGATTTTATGACGGGTACGACGCAATGCCATC	2820
2821	y		AAITTCACCTACAACATATCCATTTCTGGGCGGTAAATGTCCACCTTGGTGAGCAAAACTCA	2880
2821	b		AAITTCACCTACAACATATCCATTTCTGGGCGGTAAATGTCCACCTTGGTGAGCAAAACTCA	2880
2881	y		AGCAGCAGCATTTACGGGGAATATTACTATTCGAGAAAGCAGCAAAATTTACGCTAGAACCC	2940
2881	b		AGCAGCAGCATTTACGGGGAATATTACTATTCGAGAAAGCAGCAAAATTTACGCTAGAACCC	2940
2941	y		AATAACGCCCTANTACGCAAAACATAAAGGATAGAGTTATAAACTTCGCGAGCTTGCTC	3000
2941	b		AATAACGCCCTANTACGCAAAACATAAAGGATAGAGTTATAAACTTCGCGAGCTTGCTC	3000
3001	y		GTTAATGGGAGTTTAAGTTTAACTTGGCGAAAATATGACAGATTTAAAGGCATCTCACTATT	3060
3001	b		GTTAATGGGAGTTTAAGTTTAACTTGGCGAAAATATGACAGATTTAAAGGCATCTCACTATT	3060
3061	y		TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCTTAATATCACCAGCAATTTT	3120
3061	b		TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCTTAATATCACCAGCAATTTT	3120
3121	y		ACCAATAATGGCAGCTGCCGAAATTAATATAACAAGAGGTGGTAAAACTTGGCAATGTT	3180
3121	b		ACCAATAATGGCAGCTGCCGAAATTAATATAACAAGAGGTGGTAAAACTTGGCAATGTT	3180
3181	y		ACCAATGATGGTGAATTTAAACATTACCCTACGCTCAAGCGCAACCAAGAACGATCATC	3240

Db 3181 ACCAATGATGGTGAATTAACACATTACCACCTACGCTAAACGCAACCAAGAACATCATC 3240
QY 3241 GCGGAGATATATCAACAAAAAGGAAGCTTAATATTTACAGACAGTAATTAATGATGCT 3300
Db 3241 GCGGAGATATATCAACAAAAAGGAAGCTTAATATTTACAGACAGTAATTAATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAACGCAACCTTCAGATTTCTTCGGAT 3360
Db 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAACGCAACCTTCAGATTTCTTCGGAT 3360
QY 3361 AAAATTAATATACCAACACAGATACAAATCAAAAGGATTTGATGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATATACCAACACAGATACAAATCAAAAGGATTTGATGAGAGGACTCTAGT 3420
QY 3421 TCAGATGCGACAGTAATGCGCAACTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480
Db 3421 TCAGATGCGACAGTAATGCGCAACTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480
QY 3481 GACCTAAGTATTTTCAGGTTTCAATAAGCAGAGATTACAGCCAAAGATGATGATGATTTA 3540
Db 3481 GACCTAAGTATTTTCAGGTTTCAATAAGCAGAGATTACAGCCAAAGATGATGATGATTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCGGAAGCCAAAACAGTAACCTTTTAA 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCGGAAGCCAAAACAGTAACCTTTTAA 3600
QY 3601 AATGTTAAAGATTCAAAAAATCTCTGCTGACGGTCAACAATGTGACACTTAATAGCAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAAAATCTCTGCTGACGGTCAACAATGTGACACTTAATAGCAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGCGGACGTGAAAGCAATPAGCAGCAACGATACCGGCTTAAT 3720
Db 3661 AAAACATCTAGCAGCAATGCGGACGTGAAAGCAATPAGCAGCAACGATACCGGCTTAAT 3720
QY 3721 ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTACTTCTCAAAACAGTAATATTC 3780
Db 3721 ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTACTTCTCAAAACAGTAATATTC 3780
QY 3781 ACCGCTGCGAAAGGTTTACCAACACAGCAGCGCTGACCATTTAAGCAACAAATGGCAAA 3840
Db 3781 ACCGCTGCGAAAGGTTTACCAACACAGCAGCGCTGACCATTTAAGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACCAACCAACAGCTGATATACAGCGGTAGATTTCCGGTAAACACGTAAGT 3900
Db 3841 GCAAGTATTACCAACCAACAGCTGATATACAGCGGTAGATTTCCGGTAAACACGTAAGT 3900
QY 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960
Db 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960
QY 3961 GAGGCTAATGTAACAAAGTGAACAGGTACAATTTGCGGTACAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGGCTAATGTAACAAAGTGAACAGGTACAATTTGCGGTACAATTTCCGGTAAATACGGTA 4020
QY 4021 AATGTTACGCAAAACGCTGGCGATTTACAGTTGGGAATGCGCAGAAATTAATGCGACA 4080
Db 4021 AATGTTACGCAAAACGCTGGCGATTTACAGTTGGGAATGCGCAGAAATTAATGCGACA 4080
QY 4081 GAAGGAGCTGCAACCTTTAACCGCAACAGGAATACCTTTGACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTTAACCGCAACAGGAATACCTTTGACTACTGAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCACTAAGGCTCAGGTAGACCTCTTGCTCAGAAATGATGATGATGATGATGATG 4200
Db 4141 ATCACTTCACTAAGGCTCAGGTAGACCTCTTGCTCAGAAATGATGATGATGATGATGATG 4200
QY 4201 ATTAATGCTGTAATGACATTAATTAATACAGGACCTTAACACCGTGGCAGGCTCG 4260
Db 4201 ATTAATGCTGTAATGACATTAATTAATACAGGACCTTAACACCGTGGCAGGCTCG 4260
QY 4261 GATATTAAAGCAACCGGCGCCTTGGTTTAAAGCAAAAGATGCTAAGCTAAATGCT 4320
Db 4261 GATATTAAAGCAACCGGCGCCTTGGTTTAAAGCAAAAGATGCTAAGCTAAATGCT 4320

QY 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACGCAAGCGCTCTGCTAGTGTG 4380
Db 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACGCAAGCGCTCTGCTAGTGTG 4380
QY 4381 ACTGCGCAACCTCAACGAGTGTGAATATACATGCGGATTTAAACACAGTAATGGGTTA 4440
Db 4381 ACTGCGCAACCTCAACGAGTGTGAATATACATGCGGATTTAAACACAGTAATGGGTTA 4440
QY 4441 AATATCATTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATTTGAGTGTG 4500
Db 4441 AATATCATTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATTTGAGTGTG 4500
QY 4501 AATATCATTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATTTGAGTGTG 4560
Db 4501 AATATCATTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATTTGAGTGTG 4560
QY 4561 GAAAGATTAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTTAACTTTGGTGTAAAT 4620
Db 4561 GAAAGATTAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTTAACTTTGGTGTAAAT 4620
QY 4621 GCTGTACGTTTGTGTTAGCCAAATTAATACAAATTAACAGTCAATACACAAATTAATTAACA 4680
Db 4621 GCTGTACGTTTGTGTTAGCCAAATTAATACAAATTAACAGTCAATACACAAATTAATTAACA 4680
QY 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAGGTAAGCGCTGTTCTCAAGTGTAAAT 4740
Db 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAGGTAAGCGCTGTTCTCAAGTGTAAAT 4740
QY 4741 GCGGCACAGTATGTACCAATTTGCTGACGATGACGAGCGGTAGTCAAGTGTAAAT 4800
Db 4741 GCGGCACAGTATGTACCAATTTGCTGACGATGACGAGCGGTAGTCAAGTGTAAAT 4800
QY 4801 GTAGATTTCACTCCGCAATGAATTTATTTCTGATGAGTCAAGTGTAAATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTCACTCCGCAATGAATTTATTTCTGATGAGTCAAGTGTAAATTTACTGTGGGTTAAA 4860
QY 4861 GTTCAGTACGCGCTTTACCCATCTTTGTAATAATTTACGAGAAATTAACAGTATTTT 4920
Db 4861 GTTCAGTACGCGCTTTACCCATCTTTGTAATAATTTACGAGAAATTAACAGTATTTT 4920
QY 4921 AACAGGTTATTTATG 4937
Db 4921 AACAGGTTATTTATG 4937

RESULT 6

US-08-617-697-3

Sequence 3, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,697

FILING DATE: 01-APR-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

?	APPLICATION NUMBER:	US 08/302,832
?	FILING DATE:	05-OCT-1994
?	PRIOR APPLICATION DATA:	
?	APPLICATION NUMBER:	US PCT/US93/02166
?	FILING DATE:	16-MAR-1993
?	ATTORNEY/AGENT INFORMATION:	
?	NAME:	Berkstresser, Jerry W
?	REGISTRATION NUMBER:	22,651
?	REFERENCE/DOCKET NUMBER:	1038-557
?	TELECOMMUNICATION INFORMATION:	
?	TELEPHONE:	(703) 415-0810
?	TELEFAX:	(703) 415-0813
?	INFORMATION FOR SEQ ID NO:	3:
?	SEQUENCE CHARACTERISTICS:	
?	LENGTH:	4937 base pairs
?	TYPE:	nucleic acid
?	STRANDEDNESS:	single
?	TOPOLOGY:	linear
?	MOLECULE TYPE:	DNA (genomic)
?	US-08-617-697-3	
?	Query Match	100.0%; Score 4937; DB 2; Length 4937;
?	Best Local Similarity	100.0%; Pred. No. 0;
?	Matches 4937; Conservative	0; Mismatches 0; Indels 0; Gaps
QY	1	TAATATACAAGATAATAAAAAATAATCAAGATTTTTGTGATGACAAACAACAATTACAA 60
DB	1	TAATATACAAGATAATAAAAAATAATCAAGATTTTTGTGATGACAAACAACAATTACAA 60
QY	61	CACCTTTTTGTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAAAATCGCCATATA 120
DB	61	CACCTTTTTGTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAAAATCGCCATATA 120
QY	121	AATGGTATAATCTTTTCATCTTTTCATCTTTAACTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
DB	121	AATGGTATAATCTTTTCATCTTTTCATCTTTAACTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
QY	181	CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
DB	181	CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
QY	241	CACATGAATATGATGAACCGAGGGAAGGGAGGGGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG 300
DB	241	CACATGAATATGATGAACCGAGGGAAGGGAGGGGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG 300
QY	301	GAACGCAAAATGATAAAGTAATTAATTTGTTTCAACTTACCTTTAGGAGAAATAATGTAACCAAG 360
DB	301	GAACGCAAAATGATAAAGTAATTAATTTGTTTCAACTTACCTTTAGGAGAAATAATGTAACCAAG 360
QY	361	ATATATCGTCTCAAAATTCAGCAAAACGCCTGAATTCGTTTGGTTGCTGTGCTGAAATGGCA 420
DB	361	ATATATCGTCTCAAAATTCAGCAAAACGCCTGAATTCGTTTGGTTGCTGTGCTGAAATGGCA 420
QY	421	CGGGTGTGTGACCAATTCACAGAAAAAGGCTTCGGCTATGTTACTATCTTTTAGTGTATAC 480
DB	421	CGGGTGTGTGACCAATTCACAGAAAAAGGCTTCGGCTATGTTACTATCTTTTAGTGTATAC 480
QY	481	CACCTTAGCGTTTAAAGCCACTTTCCGCGTATGTTTACTATCTTTTAGGTTGAACATCTATTCCA 540
DB	481	CACCTTAGCGTTTAAAGCCACTTTCCGCGTATGTTTACTATCTTTTAGGTTGAACATCTATTCCA 540
QY	541	CAATCTGTTTTCAGAACGGCTTACAGGAATGGATGTAGTACACGGCACAGCCACTATG 600
DB	541	CAATCTGTTTTCAGAACGGCTTACAGGAATGGATGTAGTACACGGCACAGCCACTATG 600
QY	601	CAAGTAGATGGTATAAACCATTATCCGCAACAGTGTTCACCGTATCATTTATTCGAA 660
DB	601	CAAGTAGATGGTATAAACCATTATCCGCAACAGTGTTCACCGTATCATTTATTCGAA 660
QY	661	CAATTTAATCATCGACCAAAATGAAATGGTCAGTGTTCAGAGATTTTACAGAAAAACAACACTCCGCC 720
DB	661	CAATTTAATCATCGACCAAAATGAAATGGTCAGTGTTCAGAGATTTTACAGAAAAACAACACTCCGCC 720

Db 1801 ACCAATACAACATATTTCAAAATATCTGAAAACGCTCGACAAATGAATATACGGCAATCA 1860
QY 1861 AGAAAACCTTACCGTTAAATAGCTCAATCAACATCGGAAGCACTCCACCTTAATATCTCCAT 1920
Db 1861 AGAAAACCTTACCGTTAAATAGCTCAATCAACATCGGAAGCACTCCACCTTAATATCTCCAT 1920
QY 1921 AGTAAAGCTCAGCGTGGGAGCGGTTTCAGATGTATGAGATATTAATCTTAAGGGCGA 1980
Db 1921 AGTAAAGCTCAGCGTGGGAGCGGTTTCAGATGTATGAGATATTAATCTTAAGGGCGA 1980
QY 1981 AATTTAAACCAATTAATCTCGCGGATGGTGTGATTTATTAATAATATTAACGCTTGATCAG 2040
Db 1981 AATTTAAACCAATTAATCTCGCGGATGGTGTGATTTATTAATAATATTAACGCTTGATCAG 2040
QY 2041 GGTTTTTTAAATATTAACGCGCGTTCCTGAGCTTTTGAAGGTGGAATAAACAAGACGCG 2100
Db 2041 GGTTTTTTAAATATTAACGCGCGTTCCTGAGCTTTTGAAGGTGGAATAAACAAGACGCG 2100
QY 2101 GACGCGCAAAATGCTTAAATTTGTCGCCAGGCACTGTAAACCAATTAACAGAGGGA 2160
Db 2101 GACGCGCAAAATGCTTAAATTTGTCGCCAGGCACTGTAAACCAATTAACAGAGGGA 2160
QY 2161 GATTTACAGGCTTAAACAGCTATCTTTAAACGGAACGGTAAAGGTCTGAATATCATTTCA 2220
Db 2161 GATTTACAGGCTTAAACAGCTATCTTTAAACGGAACGGTAAAGGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280
Db 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAAACTACGAGAAAGACACCTCTGATTTGGCAACACAGGCTGATTCGACTGG 2340
Db 2281 ATTAACCAAACTACGAGAAAGACACCTCTGATTTGGCAACACAGGCTGATTCGACTGG 2340
QY 2341 AAGCTCAGTGTCTTAATCTAGAGACAGGCGCAATTTACCTTTTAAATACATTTCA 2400
Db 2341 AAGCTCAGTGTCTTAATCTAGAGACAGGCGCAATTTACCTTTTAAATACATTTCA 2400
QY 2401 AGCAATAGCAAAAGCTTAAACACACAGTATAGAAGCTCTGAGAGGCTCAATTTAACGGC 2460
Db 2401 AGCAATAGCAAAAGCTTAAACACACAGTATAGAAGCTCTGAGAGGCTCAATTTAACGGC 2460
QY 2461 GTAATGCAACATGTCATCTCAATCTCAAGAGAGGAGGAAATTAATTTCAAAATTA 2520
Db 2461 GTAATGCAACATGTCATCTCAATCTCAAGAGAGGAGGAAATTAATTTCAAAATTA 2520
QY 2521 CCAACAGAACATGAACACAAAGCACTTTACCAATTCGGTTTTAGCCAATATCA 2580
Db 2521 CCAACAGAACATGAACACAAAGCACTTTACCAATTCGGTTTTAGCCAATATCA 2580
QY 2581 GCCACTGTGGGGCTCTGTTTTTTTTGATATATATGCAACCACTTTAGCCAATATCA 2640
Db 2581 GCCACTGTGGGGCTCTGTTTTTTTTGATATATATGCAACCACTTTAGCCAATATCA 2640
QY 2641 GAGTTAAATAGTGAATTAATATCTTAACGGCGCTTAATTTTACCTTTAAATTCCTCAT 2700
Db 2641 GAGTTAAATAGTGAATTAATATCTTAACGGCGCTTAATTTTACCTTTAAATTCCTCAT 2700
QY 2701 GTTTCGGCGCATGACGCTTTTAAATCAACAAGACTTTAACCAATAATGCAACCAATTTCA 2760
Db 2701 GTTTCGGCGCATGACGCTTTTAAATCAACAAGACTTTAACCAATAATGCAACCAATTTCA 2760
QY 2761 AATTTACGCTCAGACAGCAAGAAATGATTTTTATGAGGGTACGAGCAATGCCATC 2820
Db 2761 AATTTACGCTCAGACAGCAAGAAATGATTTTTATGAGGGTACGAGCAATGCCATC 2820
QY 2821 AATTTACACCTTACACATATCCATTTGGGGGTATGTCACCCCTTGGTGCAAACTCA 2880
Db 2821 AATTTACACCTTACACATATCCATTTGGGGGTATGTCACCCCTTGGTGCAAACTCA 2880
QY 2881 AGAGAGCAATTAACGGGAATPATTACTATCGAAGAGCAGCAATGTTACGCTAGAAGCC 2940
Db 2881 AGAGAGCAATTAACGGGAATPATTACTATCGAAGAGCAGCAATGTTACGCTAGAAGCC 2940

Db 2881 AGCAGCAGCATTTACGGGGAATATTAATCTATCGAAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940
QY 2941 AATAACGCCCTTAATTCAGCAAAACATTAAGGATAGAGTTATAAACTTGGCAGCTTGCTC 3000
Db 2941 AATAACGCCCTTAATTCAGCAAAACATTAAGGATAGAGTTATAAACTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTAAAGTTTAACTGGCGAAAATGCAGATATTAAGGCAATCTCACATTT 3060
Db 3001 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAAATGCAGATATTAAGGCAATCTCACATTT 3060
QY 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCACCGCAATTTT 3120
Db 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCACCGCAATTTT 3120
QY 3121 ACCAATTAATGGCAGCTGCGGAAATTAATTAACAACAAGAGTGGTAAAACTTTGGCAATGTT 3180
Db 3121 ACCAATTAATGGCAGCTGCGGAAATTAATTAACAACAAGAGTGGTAAAACTTTGGCAATGTT 3180
QY 3181 ACCAATTAATGGTGTATTTAAACATTAACCACTCACGCTAAACGCAACCAAGAGCATCATC 3240
Db 3181 ACCAATTAATGGTGTATTTAAACATTAACCACTCACGCTAAACGCAACCAAGAGCATCATC 3240
QY 3241 GCGGAGATATAATCAACAACAAAGAGCTTAATATTAACAGACAGTAATATATGATGCT 3300
Db 3241 GCGGAGATATAATCAACAACAAAGAGCTTAATATTAACAGACAGTAATATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTCCGAT 3360
QY 3361 AAAATTAATATCAACAAACAGATAACATCAAAAAGGCTTATGATGAGAGGACTCTAGT 3420
Db 3361 AAAATTAATATCAACAAACAGATAACATCAAAAAGGCTTATGATGAGAGGACTCTAGT 3420
QY 3421 TCAGATGGCAAGTAAATGCCAACCTTAATTAACCAACAAAGAAATTAATTAATGACAGAA 3480
Db 3421 TCAGATGGCAAGTAAATGCCAACCTTAATTAACCAACAAAGAAATTAATTAATGACAGAA 3480
QY 3481 GACCTTAAGTATTTTCAAGTTTCAATTAAGCAGAGATTTACAGCAAGAGTGGTAGAGATTTA 3540
Db 3481 GACCTTAAGTATTTTCAAGTTTCAATTAAGCAGAGATTTACAGCAAGAGTGGTAGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAAATGCGGTTAAACAGCGGTGCGCAAGCGCAACAGTAACCTTTAAC 3600
Db 3541 ACTATTGGCAACAGTAAATGCGGTTAAACAGCGGTGCGCAAGCGCAACAGTAACCTTTAAC 3600
QY 3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCAATGTGACACTTAATAGCAAGTG 3660
Db 3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCAATGTGACACTTAATAGCAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAGCAATAGCGACAGTACCGGCTTAAC 3720
Db 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAGCAATAGCGACAGTACCGGCTTAAC 3720
QY 3721 ATTACTGCAAAAATGTAAGAGTAAACAAGATATTAATCTCTCAAAAACAGTAATATC 3780
Db 3721 ATTACTGCAAAAATGTAAGAGTAAACAAGATATTAATCTCTCAAAAACAGTAATATC 3780
QY 3781 ACCCGCTGCGAAGAGTTTACCACACAGCGGCTGACCATTAACGCAACAATGGCA 3840
Db 3781 ACCCGCTGCGAAGAGTTTACCACACAGCGGCTGACCATTAACGCAACAATGGCA 3840
QY 3841 GCAAGTATTAACAACCAAGAGTATACGCGGTACGATTTCCGCTAACACGTAAGT 3900
Db 3841 GCAAGTATTAACAACCAAGAGTATACGCGGTACGATTTCCGCTAACACGTAAGT 3900
QY 3901 GTTAGCGGACTGTTGATTTTAAACCACTTAATCCGCTCAAAAATTAACGCGGAAATCGGGT 3960
Db 3901 GTTAGCGGACTGTTGATTTTAAACCACTTAATCCGCTCAAAAATTAACGCGGAAATCGGGT 3960
QY 3961 GAGGCTTAATGTAACAGTGAACAGGTGACAGTACAATTTGGCGGTACAATTTCCGGTAAATACGGA 4020
Db 3961 GAGGCTTAATGTAACAGTGAACAGGTGACAGTACAATTTGGCGGTACAATTTCCGGTAAATACGGA 4020

```

QY 4021 AATGTTACGGCAAAACGCTGGGATTTAACAGTTGGGAATGCGCAGAAATTAATGCGACA 4080
DB 4021 AATGTTACGGCAAAACGCTGGGATTTAACAGTTGGGAATGCGCAGAAATTAATGCGACA 4080
QY 4081 GAAGAGCTGCAACCTTAAACGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
DB 4081 GAAGAGCTGCAACCTTAAACGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAAGGCTCAGTACAGCTCTTGCTGCTCAGATGCTAGCATGCGAGGAGC 4200
DB 4141 ATCACTTCAACTAAAGGCTCAGTACAGCTCTTGCTGCTCAGATGCTAGCATGCGAGGAGC 4200
QY 4201 ATTAATGCTGCTAAATGTGACATTAATACTACAGGCACCTTAAACACCGTGGCAGGCTCG 4260
DB 4201 ATTAATGCTGCTAAATGTGACATTAATACTACAGGCACCTTAAACACCGTGGCAGGCTCG 4260
QY 4261 GATATTAAAGCAACAGCGGACCTTGGTTATTACGCAAAAGATGCTAAAGCTAAATGGT 4320
DB 4261 GATATTAAAGCAACAGCGGACCTTGGTTATTACGCAAAAGATGCTAAAGCTAAATGGT 4320
QY 4321 GATGCAATCAGGTGATAGTACAGAAGTGAATGAGTCAAGTCACGCAAGCGGCTCTGGTAGTGTG 4380
DB 4321 GATGCAATCAGGTGATAGTACAGAAGTGAATGAGTCAAGTCACGCAAGCGGCTCTGGTAGTGTG 4380
QY 4381 ACTGCGCAACCTTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGGTTA 4440
DB 4381 ACTGCGCAACCTTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGGTTA 4440
QY 4441 AATATCATTTTCGAAAGATGTTAGAAACACTGTGCGCTTAAAGAGGCAAGGAAATGAGGTG 4500
DB 4441 AATATCATTTTCGAAAGATGTTAGAAACACTGTGCGCTTAAAGAGGCAAGGAAATGAGGTG 4500
QY 4501 AATATATTCAGGCAGGTGTAGCAAGTGTAGAAAGAGTAAATGAAAGCGGAAACGCTCCTT 4560
DB 4501 AATATATTCAGGCAGGTGTAGCAAGTGTAGAAAGAGTAAATGAAAGCGGAAACGCTCCTT 4560
QY 4561 GAAAGCTAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTAAAGCTGCTGTAAGT 4620
DB 4561 GAAAGCTAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTAAAGCTGCTGTAAGT 4620
QY 4621 GCTGAGCTTTGTTGAGCAAAATATACAAATTAACAGTCAATACACAAATGAATTTACA 4680
DB 4621 GCTGAGCTTTGTTGAGCAAAATATACAAATTAACAGTCAATACACAAATGAATTTACA 4680
QY 4681 ACCAGACGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 4740
DB 4681 ACCAGACGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 4740
QY 4741 GCGGACAGATGATGATACCAATGTTGCTGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4800
DB 4741 GCGGACAGATGATGATACCAATGTTGCTGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4800
QY 4801 GTAGATTTTCATCCTGCAATGAAGTCAATTTTATTTTCGTTATTTATTTCTGTTGGGTTAA 4860
DB 4801 GTAGATTTTCATCCTGCAATGAAGTCAATTTTATTTTCGTTATTTATTTCTGTTGGGTTAA 4860
QY 4861 GTTCAGTACGGGCTTTACCCATCTGTTGAAAAAATTTACGGAGATACAAATGAAGTATTTT 4920
DB 4861 GTTCAGTACGGGCTTTACCCATCTGTTGAAAAAATTTACGGAGATACAAATGAAGTATTTT 4920
QY 4921 AACAGGTTATTTATG 4937
DB 4921 AACAGGTTATTTATG 4937

```

```

RESULT 7
US-08-719-641-3
; Sequence 3, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins

```

```

; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berks Burgess, Terry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-719-641-3

```

```

Query Match 100.0%; Score 4937; DB 4; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATATACAAGATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 60
DB 1 TAAATATACAAGATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 60
QY 61 CACCTTTTTCGACTCTATATGCAAAATATTTTAAAAAATAAGTATAAATCCGCGCATATAA 120
DB 61 CACCTTTTTCGACTCTATATGCAAAATATTTTAAAAAATAAGTATAAATCCGCGCATATAA 120
QY 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 180
DB 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 180
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
DB 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
DB 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
QY 301 GAACGCAAAATGATAAAGTAATTTAATTTGTTCAACTAACCTTTAGGAGAAAAATATGAACAAG 360
DB 301 GAACGCAAAATGATAAAGTAATTTAATTTGTTCAACTAACCTTTAGGAGAAAAATATGAACAAG 360
QY 361 ATATATCGTCTCAAAATTCAGCAAAACGCGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 420

```

|||||
Db 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGGTTGCTGTGCTGAATGGCA 420
QY 421 CGGGGTTGTGACCAATTCACAGAAAAAGGCTTCCGCTATGTTACTATCTTTAGTGTAAAC 480
Db 421 CGGGGTTGTGACCAATTCACAGAAAAAGGCTTCCGCTATGTTACTATCTTTAGTGTAAAC 480
QY 481 CACTTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGTAAACATCTATTCCA 540
Db 481 CACTTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGTAAACATCTATTCCA 540
QY 541 CAATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGTAGTACACAGGCAAGCCACTATG 600
Db 541 CAATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGTAGTACACAGGCAAGCCACTATG 600
QY 601 CAAGTAGATGGTAAATAAACCAATTCACCAACAGTGTGACGCTATCAATTAATTTGAAA 660
Db 601 CAAGTAGATGGTAAATAAACCAATTCACCAACAGTGTGACGCTATCAATTAATTTGAAA 660
QY 661 CAATTTAATCATCGCAAAATGAATGAAATGAGTGTGCAAGTGTGTTTACAAGAAAAACAACCTCCGCC 720
Db 661 CAATTTAATCATCGCAAAATGAATGAAATGAGTGTGCAAGTGTGTTTACAAGAAAAACAACCTCCGCC 720
QY 721 GTATTTCAACCGTGTATACATCTAACCAAAATCTCCCAATTAAGGGATTTAGATTTCTAAC 780
Db 721 GTATTTCAACCGTGTATACATCTAACCAAAATCTCCCAATTAAGGGATTTAGATTTCTAAC 780
QY 781 GGACAGTCTTTTTAATCAACCAAAATGGTATACAAATAGTAAAGAGCAATTAATTAAC 840
Db 781 GGACAGTCTTTTTAATCAACCAAAATGGTATACAAATAGTAAAGAGCAATTAATTAAC 840
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGGCGGTAAT 900
Db 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGGCGGTAAT 900
QY 901 TACACCTTCGAGCAAAACCAAGATAAAGCGTCTGCTGAAATTTGTAATCAGGTTAAAT 960
Db 901 TACACCTTCGAGCAAAACCAAGATAAAGCGTCTGCTGAAATTTGTAATCAGGTTAAAT 960
QY 961 ACTGTCGTAAGACGCGAGTGTAAATCTTATTGTTGGCAAAATGCAAAACGAGGCTGTG 1020
Db 961 ACTGTCGTAAGACGCGAGTGTAAATCTTATTGTTGGCAAAATGCAAAACGAGGCTGTG 1020
QY 1021 ATTACGTAATTTGCGAGCAATTTCTTACTCGAGGCAAAACCAATCAACCATCAGCAT 1080
Db 1021 ATTACGTAATTTGCGAGCAATTTCTTACTCGAGGCAAAACCAATCAACCATCAGCAT 1080
QY 1081 ATAATAAACCAACCAATTTACTTACAGCATTTCCGCGCTGAAATGAAAGCGGTCATCTG 1140
Db 1081 ATAATAAACCAACCAATTTACTTACAGCATTTCCGCGCTGAAATGAAAGCGGTCATCTG 1140
QY 1141 GCGATATTTTCCCAAGCGGTAACATTAATGTCGCGTGCCTGCACTATTGCAACCAA 1200
Db 1141 GCGATATTTTCCCAAGCGGTAACATTAATGTCGCGTGCCTGCACTATTGCAACCAA 1200
QY 1201 GGTAACTTTCTGCTGATTTGTAAGCAAAAGATAAAGGCGCAATATTGTTCTTTCCGCC 1260
Db 1201 GGTAACTTTCTGCTGATTTGTAAGCAAAAGATAAAGGCGCAATATTGTTCTTTCCGCC 1260
QY 1261 AAAGAGGTGAAGCGGAATTTGCGGTTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
Db 1261 AAAGAGGTGAAGCGGAATTTGCGGTTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
QY 1321 GCAAGCTGATGATTACAGGCGATAAAGTACATTAATAACAGGTGCGAGTTATCGACCTT 1380
Db 1321 GCAAGCTGATGATTACAGGCGATAAAGTACATTAATAACAGGTGCGAGTTATCGACCTT 1380
QY 1381 TCAGGTAAGAGGGGGAGAACTTACCTTGGCGGTGACAGCGCGGCAAGGTAAAC 1440
Db 1381 TCAGGTAAGAGGGGGAGAACTTACCTTGGCGGTGACAGCGCGGCAAGGTAAAC 1440
QY 1441 GCGATTCATTTAGCAAGAAACCTTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500
|||||

Db 1441 GGCATTCAATTTAGCAAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500
QY 1501 AAGAAAAAGGCGGACGCGCTATTGTTGGGCGGATATTGCGTTAAATTCACGCGCAATATT 1560
Db 1501 AAGAAAAAGGCGGACGCGCTATTGTTGGGCGGATATTGCGTTAAATTCACGCGCAATATT 1560
QY 1561 AACGCTCAAGGTAGTGTGATATCGCTAAACCCGGTGGTTTTGTTGGAGACATCGGGGCAT 1620
Db 1561 AACGCTCAAGGTAGTGTGATATCGCTAAACCCGGTGGTTTTGTTGGAGACATCGGGGCAT 1620
QY 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAAAAAGAGTGTGTTAGACCCCTGAT 1680
Db 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAAAAAGAGTGTGTTAGACCCCTGAT 1680
QY 1681 GATGTAACTTTGAGCGGAGACCCCTTCGCAATTAATACCGGTATTAATGATGAATTC 1740
Db 1681 GATGTAACTTTGAGCGGAGACCCCTTCGCAATTAATACCGGTATTAATGATGAATTC 1740
QY 1741 CCAACAGGCGGCGTGAAGCAAGCGACCTAAAAAATAGCGAACTCAAAACCAAGCCTA 1800
Db 1741 CCAACAGGCGGCGTGAAGCAAGCGACCTAAAAAATAGCGAACTCAAAACCAAGCCTA 1800
QY 1801 ACCAATACAACTATTTCCTCAATTTATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1860
Db 1801 ACCAATACAACTATTTCCTCAATTTATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1860
QY 1861 AGAAACTTACCGTTTAAAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920
Db 1861 AGAAACTTACCGTTTAAAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920
QY 1921 AGTAAAGTACGCGTGGCGGCGTTCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980
Db 1921 AGTAAAGTACGCGTGGCGGCGTTCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980
QY 1981 AATTTAACTTTTATTTCTGCGGATGGTTCATGTTCTATAAAATATTACGCTTGATCAG 2040
Db 1981 AATTTAACTTTTATTTCTGCGGATGGTTCATGTTCTATAAAATATTACGCTTGATCAG 2040
QY 2041 GGTTTTTTAAATATTACCCTCCCTTCCGTAGCTTTTGAAGGTGGAATTAACAAGCAAGC 2100
Db 2041 GGTTTTTTAAATATTACCCTCCCTTCCGTAGCTTTTGAAGGTGGAATTAACAAGCAAGC 2100
QY 2101 GACGCGCAATGCTAAATTTCTGCGGCGGCGTTCAGATTTGATGTTCAATAAAATATTACGCTTGATCAG 2160
Db 2101 GACGCGCAATGCTAAATTTCTGCGGCGGCGTTCAGATTTGATGTTCAATAAAATATTACGCTTGATCAG 2160
QY 2161 GATTTACGGGCTAACACGTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2220
Db 2161 GATTTACGGGCTAACACGTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTTAACCCCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280
Db 2221 TCAGTGAATTAATTTAACCCCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAAACCAACTACGAGAAAGAACACCTCGTATTGGCAAAACGCGCATGATTCGCACTGG 2340
Db 2281 ATTAAACCAACTACGAGAAAGAACACACCTCGTATTGGCAAAACGCGCATGATTCGCACTGG 2340
QY 2341 AACGCTAGTCTCTTAAATCTAGACAGCGCAATTTTACCTTTTAAATACATTTCA 2400
Db 2341 AACGCTAGTCTCTTAAATCTAGACAGCGCAATTTTACCTTTTAAATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGCGGGGTGAATTTTAAAGGC 2460
Db 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGCGGGGTGAATTTTAAAGGC 2460
QY 2461 GATAATGGCAACATGCTCAATTTCAATCTCAAGAGGCGGAAGCTTAATTTCAATTTAAA 2520
Db 2461 GATAATGGCAACATGCTCAATTTCAATCTCAAGAGGCGGAAGCTTAATTTCAATTTAAA 2520
QY 2521 CCAACGCAACATGACACAGCAACCTTTTACCAATTCGTTTTTACCAATATCACA 2580
Db 2521 CCAACGCAACATGACACAGCAACCTTTTACCAATTCGTTTTTACCAATATCACA 2580

Tue Mar 25 08:39:01 2003

QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTTCATATATATGCGCAACCAATTCCTGGCAGAGGGCT 2640
DB 2581 GCCACTGGTGGGGCTCTGTTTTTTTTCATATATATGCGCAACCAATTCCTGGCAGAGGGCT 2640
QY 2641 GAGTTTAAAAATGAGTGAATTAATATATCTAAGCGCGCTAAATTTTACCTTAAATTCGCAT 2700
DB 2641 GAGTTTAAAAATGAGTGAATTAATATATCTAAGCGCGCTAAATTTTACCTTAAATTCGCAT 2700
QY 2701 GTTTCGGCGGATGAGCGCTTTTAAAAATCAACAAAGACTTTAAACCATAAATGCAACCAATFCA 2760
DB 2701 GTTTCGGCGGATGAGCGCTTTTAAAAATCAACAAAGACTTTAAACCATAAATGCAACCAATFCA 2760
QY 2761 AATTTTCAGCCTCAGACAGACGAAAGATGATTTTTATGAGCGGTACGCGCAATGCCATC 2820
DB 2761 AATTTTCAGCCTCAGACAGACGAAAGATGATTTTTATGAGCGGTACGCGCAATGCCATC 2820
QY 2821 AATTTCAACCTTACCAATATCCATCTTGGCGGTAATGTCAACCTTTGGTGGAACCAATCA 2880
DB 2821 AATTTCAACCTTACCAATATCCATCTTGGCGGTAATGTCAACCTTTGGTGGAACCAATCA 2880
QY 2881 ACCAGCAGCATTTACGGGGAATTTACTATCGAGAAAGACGAAATCTTACGCTAGAAGCC 2940
DB 2881 ACCAGCAGCATTTACGGGGAATTTACTATCGAGAAAGACGAAATCTTACGCTAGAAGCC 2940
QY 2941 AATAAGCGCCCTAATCAGCAAAACATTAAGGATAGAGTTTAAACTTTGGCAGCTTGCTC 3000
DB 2941 AATAAGCGCCCTAATCAGCAAAACATTAAGGATAGAGTTTAAACTTTGGCAGCTTGCTC 3000
QY 3001 GTTAATGGGAGTTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTATT 3060
DB 3001 GTTAATGGGAGTTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCACCGCAATTTT 3120
DB 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCACCGCAATTTT 3120
QY 3121 ACCAATATGCGACTGCCGAAATTAATATTAACACAGGAGTGTGTAACCTTGCAATGTT 3180
DB 3121 ACCAATATGCGACTGCCGAAATTAATATTAACACAGGAGTGTGTAACCTTGCAATGTT 3180
QY 3181 ACCAATGATGGTATTTAAACATTTACCACCTCACGCTTAAACGCAACCAAGACATCATC 3240
DB 3181 ACCAATGATGGTATTTAAACATTTACCACCTCACGCTTAAACGCAACCAAGACATCATC 3240
QY 3241 GCGGAGATATAATCAACAAAAAGAGGCTTAAATATTAACAGACATTAATATGATGCT 3300
DB 3241 GCGGAGATATAATCAACAAAAAGAGGCTTAAATATTAACAGACATTAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTCACGATTTCTTCGGAT 3360
DB 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTCACGATTTCTTCGGAT 3360
QY 3361 AAAATTAATATCAACACAGATTAACATCAAAAGGCTATTTGATGGAGGAGCTCTAGT 3420
DB 3361 AAAATTAATATCAACACAGATTAACATCAAAAGGCTATTTGATGGAGGAGCTCTAGT 3420
QY 3421 TCAGATGGACAGTAATGCAACCTTAATTAACCAAGCAATTTGAAATTCACAGAA 3480
DB 3421 TCAGATGGACAGTAATGCAACCTTAATTAACCAAGCAATTTGAAATTCACAGAA 3480
QY 3481 GACCTAAGTATTTAGGTTTCAATTAACAGAGATTAACAGCAAGATGTTAGAGATTTA 3540
DB 3481 GACCTAAGTATTTAGGTTTCAATTAACAGAGATTAACAGCAAGATGTTAGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCGGAGCCCAACCAAGTAACTTTAAC 3600
DB 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCGGAGCCCAACCAAGTAACTTTAAC 3600
QY 3601 AATGTTTAAAGATTCAAAAATCTCTGCTGACGGTCAACAATGTCACACTAAATAGCAAGTG 3660
DB 3601 AATGTTTAAAGATTCAAAAATCTCTGCTGACGGTCAACAATGTCACACTAAATAGCAAGTG 3660

QY 3661 AAAACATCTAGCAGCAATGCGGACGTGAAGCAATAGCGCAACGATACCGGCTTAACT 3720
DB 3661 AAAACATCTAGCAGCAATGCGGACGTGAAGCAATAGCGCAACGATACCGGCTTAACT 3720
QY 3721 ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAAACAGTAAATATC 3780
DB 3721 ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAAACAGTAAATATC 3780
QY 3781 ACCGCTCGGAAAGGTTTACCACACAGCAGGCTCGACCATTAACGCAACAAATGCGCAA 3840
DB 3781 ACCGCTCGGAAAGGTTTACCACACAGCAGGCTCGACCATTAACGCAACAAATGCGCAA 3840
QY 3841 GCAAGTATTACAACCAACAAAGGTGATACAGCGGTACGATTTCCGGTACACAGGTAAGT 3900
DB 3841 GCAAGTATTACAACCAACAAAGGTGATACAGCGGTACGATTTCCGGTACACAGGTAAGT 3900
QY 3901 GTTAGCGGACTGTGATTTTAAACCACTTAATCCGCTCAAAAATTTGAAGCAAAATCCGGT 3960
DB 3901 GTTAGCGGACTGTGATTTTAAACCACTTAATCCGCTCAAAAATTTGAAGCAAAATCCGGT 3960
QY 3961 GAGGCTAATGTAAAGTGAACAGGTACAAATTCGCGGTACAAATTCGCGTAAATACGTA 4020
DB 3961 GAGGCTAATGTAAAGTGAACAGGTACAAATTCGCGGTACAAATTCGCGTAAATACGTA 4020
QY 4021 AATGTTACGGCAACGCTGGCGATTTTAAACAGTTGGGAATGCGGCAAAATTTAATCCGCA 4080
DB 4021 AATGTTACGGCAACGCTGGCGATTTTAAACAGTTGGGAATGCGGCAAAATTTAATCCGCA 4080
QY 4081 GAAAGGCTGCAACCTTAAACCGCAACAGGAAATACCTTACTACTGAAAGCGGTTCTAGC 4140
DB 4081 GAAAGGCTGCAACCTTAAACCGCAACAGGAAATACCTTACTACTGAAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTTAAGGCTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCGAGGAAGC 4200
DB 4141 ATCACTTCAACTTAAGGCTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCGAGGAAGC 4200
QY 4201 ATTAATGCTGCTAAATGTGACATTAATACACAGCACCTTAAACCAACGCTGCGAGGCTG 4260
DB 4201 ATTAATGCTGCTAAATGTGACATTAATACACAGCACCTTAAACCAACGCTGCGAGGCTG 4260
QY 4261 GATATTAAAGCAACAGCGGCACTTTGGTTATTAAACGCAAAAGATGCTAAAGTAAATGGT 4320
DB 4261 GATATTAAAGCAACAGCGGCACTTTGGTTATTAAACGCAAAAGATGCTAAAGTAAATGGT 4320
QY 4321 GATGCTCAGGTGATAGTACAGAAAGTAAATGCAAGTCAACGCAAGCGGCTCTGCTAGTGTG 4380
DB 4321 GATGCTCAGGTGATAGTACAGAAAGTAAATGCAAGTCAACGCAAGCGGCTCTGCTAGTGTG 4380
QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCCTTGGGATTTAAACACAGTAAATGGTTA 4440
DB 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCCTTGGGATTTAAACACAGTAAATGGTTA 4440
QY 4441 AATATCATTTTGAAGATGGTAGAACACTGTGCGCTTAAGAGCAAGAAATTTAGGTG 4500
DB 4441 AATATCATTTTGAAGATGGTAGAACACTGTGCGCTTAAGAGCAAGAAATTTAGGTG 4500
QY 4501 AATATATATCAGCAGGCTGAGCAAGTGTAGAAGAGTAAATTTGAAGCAAGCAAGGCTCTT 4560
DB 4501 AATATATATCAGCAGGCTGAGCAAGTGTAGAAGAGTAAATTTGAAGCAAGCAAGGCTCTT 4560
QY 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTTAACTTTGGTGAAGT 4620
DB 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTTAACTTTGGTGAAGT 4620
QY 4621 GCTGTACGTTTGTGAGCCAAATTAATACAAATTTACAGTCAATACACAAAATGAATTTACA 4680
DB 4621 GCTGTACGTTTGTGAGCCAAATTAATACAAATTTACAGTCAATACACAAAATGAATTTACA 4680
QY 4681 ACCAGCGCTCAAGTCAAGTGAATTTCTGAAGGTGAAGCGGTGTTTCTCAAGTGGTAAAT 4740
DB 4681 ACCAGCGCTCAAGTCAAGTGAATTTCTGAAGGTGAAGCGGTGTTTCTCAAGTGGTAAAT 4740
QY 4741 GCGCAGCAGTATGTACCAATTTGCTGACGATGGACAGCGGTAGTCAGTAAATTTGCAAG 4800

Db 4741 GCGCACAGTATGACCAATGTTCTGACGATGGACGCCGTAGTCAGTAATGACAAG 4800
QY 4801 GTAGATTTCAATCCGCAATGAAGTCAATTTATTTTCTGATTTACTGTTGGGTTAAA 4860
Db 4801 GTAGATTTCAATCCGCAATGAAGTCAATTTATTTTCTGATTTACTGTTGGGTTAAA 4860
QY 4861 GTTCAAGTACGGGCTTTACCCATCTGTGTAATAAATACGGAGAATACAAATGAATTTT 4920
Db 4861 GTTCAAGTACGGGCTTTACCCATCTGTGTAATAAATACGGAGAATACAAATGAATTTT 4920
QY 4921 AACAGGTTATTTATG 4937
Db 4921 AACAGGTTATTTATG 4937

RESULT 8

US-09-206-942-70
: Sequence 70, Application US/09206942
: Patent No. 6432669
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-Ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
: FILE OF INVENTION: Molecular Weight Proteins
: FILE REFERENCE: 1038-861 MTS:jb
: CURRENT APPLICATION NUMBER: US/09/206.942
: EARLIER FILING DATE: 1998-12-08
: EARLIER APPLICATION NUMBER: 09/167.568
: EARLIER FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 70
: LENGTH: 4937
: TYPE: DNA
: ORGANISM: Haemophilus influenzae
US-09-206-942-70

Query Match 100.0%; Score 4937; DB 4; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAAATACAA 60
Db 1 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAAATACAA 60
QY 61 CACCTTTTTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCATATA 120
Db 61 CACCTTTTTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCATATA 120
QY 121 AATGATATAATCTTTCATCTTTTCATCTTTTATCTTTTCATCTTTTCATCTTTTCAT 180
Db 121 AATGATATAATCTTTCATCTTTTCATCTTTTATCTTTTCATCTTTTCATCTTTTCAT 180
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Db 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
QY 241 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db 241 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 301 GACGCAATATGATGAAGTAATTTAATTTTCACTAACCTTTAGGAGAAATATGAACAAG 360
Db 301 GACGCAATATGATGAAGTAATTTAATTTTCACTAACCTTTAGGAGAAATATGAACAAG 360
QY 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGGTCTGCTGCTGCTGAATGGCA 420
Db 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGGTCTGCTGCTGCTGAATGGCA 420
QY 421 CGGGGTTGTGACCATTTCCACAGAAAAGGCTTCGCTATGTTACTATCTTTAGGTGTAAC 480
Db 421 CGGGGTTGTGACCATTTCCACAGAAAAGGCTTCGCTATGTTACTATCTTTAGGTGTAAC 480

Db 421 CGGGGTTGTGACCATTTCCACAGAAAAGGCTTCGCTATGTTACTATCTTTAGGTGTAAC 480
QY 481 CACTTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540
Db 481 CACTTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540
QY 541 CAATCTCTGTTTACGAAGCGGCTTACAAGGAATGGATGTAGTACACGGCAGCCACTATG 600
Db 541 CAATCTCTGTTTACGAAGCGGCTTACAAGGAATGGATGTAGTACACGGCAGCCACTATG 600
QY 601 CAAGTAGATGTAATAAAACCATTTATCCGCAACAGTGTGACGCTATCAATTAATTTGGAAA 660
Db 601 CAAGTAGATGTAATAAAACCATTTATCCGCAACAGTGTGACGCTATCAATTAATTTGGAAA 660
QY 661 CAATTTAATCATCGACCAAAATGAATGGTGGAGTTTTCACAGAAAACCAACTCCGCC 720
Db 661 CAATTTAATCATCGACCAAAATGAATGGTGGAGTTTTCACAGAAAACCAACTCCGCC 720
QY 721 GTATTCAACCGTGTTCATCTTAACCAATCTCCCAATTTAAAGGGATTTTAGATTTCTAAC 780
Db 721 GTATTCAACCGTGTTCATCTTAACCAATCTCCCAATTTAAAGGGATTTTAGATTTCTAAC 780
QY 781 GGACAAGTCTTTTAAATCAACCCAAATGATATCAATAGTAAAGCGCAATTTATTAAC 840
Db 781 GGACAAGTCTTTTAAATCAACCCAAATGATATCAATAGTAAAGCGCAATTTATTAAC 840
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 900
Db 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 900
QY 901 TTCACCTTCGAGCAACCAAGATTAAGCGCTCGCTGAAATTTGTGAATTCACGCTTTAAT 960
Db 901 TTCACCTTCGAGCAACCAAGATTAAGCGCTCGCTGAAATTTGTGAATTCACGCTTTAAT 960
QY 961 ACTGTGCGTAAAGACGGCAGTGTAAATCTTATTTGGTGGCAAGTGTAAAGAGGCTGTG 1020
Db 961 ACTGTGCGTAAAGACGGCAGTGTAAATCTTATTTGGTGGCAAGTGTAAAGAGGCTGTG 1020
QY 1021 ATTACGCTAAATGTGGCAGCATTTCTTACTCGAGGCAAAAATCACCATTCAGCGAT 1080
Db 1021 ATTACGCTAAATGTGGCAGCATTTCTTACTCGAGGCAAAAATCACCATTCAGCGAT 1080
QY 1081 ATAATAAACCCACCACTTACTTACAGCATTTCCGCGCTGAAATTTGAAGCGGTCAATCTG 1140
Db 1081 ATAATAAACCCACCACTTACTTACAGCATTTCCGCGCTGAAATTTGAAGCGGTCAATCTG 1140
QY 1141 GCGCATATTTTGGCAAGGCGGTAACTTAATGTCGCTGTCACATTTATTCGAAACCAA 1200
Db 1141 GCGCATATTTTGGCAAGGCGGTAACTTAATGTCGCTGTCACATTTATTCGAAACCAA 1200
QY 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATTAAGGGGCAATATTTCTTTCCGCC 1260
Db 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATTAAGGGGCAATATTTCTTTCCGCC 1260
QY 1261 AAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320
Db 1261 AAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320
QY 1321 GGCAAGCTGATGATTTACAGCGGATTAAGTACATTTAAAGAGGTGAGTATTCGACCTT 1380
Db 1321 GGCAAGCTGATGATTTACAGCGGATTAAGTACATTTAAAGAGGTGAGTATTCGACCTT 1380
QY 1381 TCAGGTAAGAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGCGCAAGGTAAGAAC 1440
Db 1381 TCAGGTAAGAGAGGGGAGAACTTACCTTTGGCGGTGACGAGCGCGCGCAAGGTAAGAAC 1440
QY 1441 GGCATTCATTTAGCAAGAAACCTCTTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500
Db 1441 GGCATTCATTTAGCAAGAAACCTCTTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500
QY 1501 AAAGAAAAGGCGGACGCGCTATTTGTGGGGGATATTTGCGTTAATTTACGCGCAATATT 1560
Db 1501 AAAGAAAAGGCGGACGCGCTATTTGTGGGGGATATTTGCGTTAATTTACGCGCAATATT 1560

QY 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTTTGTGGAGACATCGGGCAT 1620
DB 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTTTGTGGAGACATCGGGCAT 1620
QY 1621 TATTTATPCCATGACAGCAATCAATGTTTAAACAAAAGAGTGGTGTGACCCCTGAT 1680
DB 1621 TATTTATPCCATGACAGCAATCAATGTTTAAACAAAAGAGTGGTGTGACCCCTGAT 1680
QY 1681 GATGTAAACAATTGAAGCCGAAGACCCCTTTCGCAATAATACCCGGTATAATGATGAATTC 1740
DB 1681 GATGTAAACAATTGAAGCCGAAGACCCCTTTCGCAATAATACCCGGTATAATGATGAATTC 1740
QY 1741 CCAACAGGCACCGGTGAAGCAAGCGACCCCTTAAAAAATAAGCGAACTCAAAACAAGCCTA 1800
DB 1741 CCAACAGGCACCGGTGAAGCAAGCGACCCCTTAAAAAATAAGCGAACTCAAAACAAGCCTA 1800
QY 1801 ACCAATACAACATTTTCAAAATTTATCTGAAAAAGCCCTGGACAATGAATATAAGCGCATCA 1860
DB 1801 ACCAATACAACATTTTCAAAATTTATCTGAAAAAGCCCTGGACAATGAATATAAGCGCATCA 1860
QY 1861 AGAAACTTACCGGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAAATTCCTCAT 1920
DB 1861 AGAAACTTACCGGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAAATTCCTCAT 1920
QY 1921 AGTAAGGTACCGGTGGCGGAGCGTTCAGATTGATGGAGATATTACTTCTTAAAGCGGA 1980
DB 1921 AGTAAGGTACCGGTGGCGGAGCGTTCAGATTGATGGAGATATTACTTCTTAAAGCGGA 1980
QY 1981 AATTTAACCATTTATCTGGCGGATGGTTGATGTTTCATATAATATTAGCTTGATCAG 2040
DB 1981 AATTTAACCATTTATCTGGCGGATGGTTGATGTTTCATATAATATTAGCTTGATCAG 2040
QY 2041 GGTTTTTAAATATTACCGCCGCTTCGTTAGCTTTTGAAGTGAATATAACAAGCAGC 2100
DB 2041 GGTTTTTAAATATTACCGCCGCTTCGTTAGCTTTTGAAGTGAATATAACAAGCAGC 2100
QY 2101 GACGGCGAAATGCTAAATTTGTCCGACAGGGCACTGTAAACCATACAGGAGGGGAAA 2160
DB 2101 GACGGCGAAATGCTAAATTTGTCCGACAGGGCACTGTAAACCATACAGGAGGGGAAA 2160
QY 2161 GATTTTCAGGGCTAACACAGATATCTTTAAACGGAACGGGTAAAGTCTGAATATCATTTCA 2220
DB 2161 GATTTTCAGGGCTAACACAGATATCTTTAAACGGAACGGGTAAAGTCTGAATATCATTTCA 2220
QY 2221 TCAGTGAATTAATTAACCCACAATCTTATAGTGGCACAAATTAACATATCTGGGAATATAACA 2280
DB 2221 TCAGTGAATTAATTAACCCACAATCTTATAGTGGCACAAATTAACATATCTGGGAATATAACA 2280
QY 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACCTGG 2340
DB 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACCTGG 2340
QY 2341 AACGTCAGTGCCTTAATCTAGAGACAGCGCCAAATTTTACCTTTATTAATATACATTTCA 2400
DB 2341 AACGTCAGTGCCTTAATCTAGAGACAGCGCCAAATTTTACCTTTATTAATATACATTTCA 2400
QY 2401 AGCAATAGCAAGGCTTAAACACACAGATATAGAACCTCTGAGGGGTGAAATTTAAACGGC 2460
DB 2401 AGCAATAGCAAGGCTTAAACACACAGATATAGAACCTCTGAGGGGTGAAATTTAAACGGC 2460
QY 2461 GTAAATGGCAACATGTCATTTCAATCTCAAGAAAGAGGCGAAAGTTTAAATTTCAATTAATA 2520
DB 2461 GTAAATGGCAACATGTCATTTCAATCTCAAGAAAGAGGCGAAAGTTTAAATTTCAATTAATA 2520
QY 2521 CCAACAGGAGAACATGAACACAAAGCAACCTTTACCAATTCGGTTTTTACGCCAATATCACA 2580
DB 2521 CCAACAGGAGAACATGAACACAAAGCAACCTTTACCAATTCGGTTTTTACGCCAATATCACA 2580
QY 2581 GCCACTTGGTGGGGCTCTGTTTTTTTGTATATATATGCCCAACCATCTGGCAGAGGGCT 2640
DB 2581 GCCACTTGGTGGGGCTCTGTTTTTTTGTATATATATGCCCAACCATCTGGCAGAGGGCT 2640

QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAAATTTTACCTTAAATTCCTCAT 2700
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAAATTTTACCTTAAATTCCTCAT 2700
QY 2701 GTTCCGCGGATGACGCTTTTAAATAATCAACAAAGACTTTAACCATATAATGCAACCAATTC 2760
DB 2701 GTTCCGCGGATGACGCTTTTAAATAATCAACAAAGACTTTAACCATATAATGCAACCAATTC 2760
QY 2761 AATTTTCAGCTTCAGACAGACGAAAGATGATTTTATGAGGGTACGACGCAATGCCATC 2820
DB 2761 AATTTTCAGCTTCAGACAGACGAAAGATGATTTTATGAGGGTACGACGCAATGCCATC 2820
QY 2821 AATTTCAACCTACACATATCCATTTCTGGCGGTAAATGTCACCTTGGTGGACAAACTCA 2880
DB 2821 AATTTCAACCTACACATATCCATTTCTGGCGGTAAATGTCACCTTGGTGGACAAACTCA 2880
QY 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGAGCAAAATGTTACGCTAGAAGCC 2940
DB 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGAGCAAAATGTTACGCTAGAAGCC 2940
QY 2941 AATAAGCCCTTAATCAGCAAAACATTAAGGATAGAGTTTATAAATTTGGCAGCTTGCTC 3000
DB 2941 AATAAGCCCTTAATCAGCAAAACATTAAGGATAGAGTTTATAAATTTGGCAGCTTGCTC 3000
QY 3001 GTTAAATGGAGTTTAAAGTTTAACTGGCGAAAATGAGATATTAAGGCAATCTCACTATT 3060
DB 3001 GTTAAATGGAGTTTAAAGTTTAACTGGCGAAAATGAGATATTAAGGCAATCTCACTATT 3060
QY 3061 TCAGAAAGCCCACTTTTAAAGGAAAAGACTAGAGATACCCTAAATATCACCGGCAATTTT 3120
DB 3061 TCAGAAAGCCCACTTTTAAAGGAAAAGACTAGAGATACCCTAAATATCACCGGCAATTTT 3120
QY 3121 ACCAATATGGCATTTAAACATTTAAACATAACACAGAGAGTGGTAAAACTTGGCAATGTT 3180
DB 3121 ACCAATATGGCATTTAAACATTTAAACATAACACAGAGAGTGGTAAAACTTGGCAATGTT 3180
QY 3181 ACCAATATGGCATTTAAACATTTAAACATAACACAGAGAGTGGTAAAACTTGGCAATGTT 3180
DB 3181 ACCAATATGGCATTTAAACATTTAAACATAACACAGAGAGTGGTAAAACTTGGCAATGTT 3180
QY 3181 ACCAATATGGCATTTAAACATTTAAACATAACACAGAGAGTGGTAAAACTTGGCAATGTT 3180
DB 3181 ACCAATATGGCATTTAAACATTTAAACATAACACAGAGAGTGGTAAAACTTGGCAATGTT 3180
QY 3241 GCGGAGATATTAATCAACAAAAGAGAGTAAATTTACACACAGTAAATATGATGCT 3300
DB 3241 GCGGAGATATTAATCAACAAAAGAGAGTAAATTTACACACAGTAAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTCAGGATTTCTCCGAT 3360
DB 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGAGGCAACCTCAGGATTTCTCCGAT 3360
QY 3361 AAAATTAATATCACCAACAGATTAACAAATCAAAAAGGATTTGATGGAGAGGACTCTAGT 3420
DB 3361 AAAATTAATATCACCAACAGATTAACAAATCAAAAAGGATTTGATGGAGAGGACTCTAGT 3420
QY 3421 TCAGATCGCAACAGTAAATGCAACCTTAACCTATTAAAAACCAAGAAATTTGAAATTTGACAGAA 3480
DB 3421 TCAGATCGCAACAGTAAATGCAACCTTAACCTATTAAAAACCAAGAAATTTGAAATTTGACAGAA 3480
QY 3481 GACCTAAGTATTTTCAGGTTTCAATATAAGCAGAGATTTACAGCCAAAAGATGGTAGAGATTTA 3540
DB 3481 GACCTAAGTATTTTCAGGTTTCAATATAAGCAGAGATTTACAGCCAAAAGATGGTAGAGATTTA 3540
QY 3541 ACTATTGGCACAGTAAATGAGGTAACAGCGGTGCCCAAGCCAAAACAGTAACCTTTTAAAC 3600
DB 3541 ACTATTGGCACAGTAAATGAGGTAACAGCGGTGCCCAAGCCAAAACAGTAACCTTTTAAAC 3600
QY 3601 AATGTTAAAGATTTCAAAAATCTCTGCTCAGGTCACATGTCACACTTAAATAGCAAGTG 3660
DB 3601 AATGTTAAAGATTTCAAAAATCTCTGCTCAGGTCACATGTCACACTTAAATAGCAAGTG 3660
QY 3661 AAAACATCTACGACGAATGGCGGAGCTGGAAGCAATAGCGCAACGATACCGGCTTAAC 3720
DB 3661 AAAACATCTACGACGAATGGCGGAGCTGGAAGCAATAGCGCAACGATACCGGCTTAAC 3720
QY 3721 ATTACTGCAAAAAATGTTAGAGTAAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780

Db 3721 ATTACTGCAAAAATGTAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780
QY 3781 ACCGGCTGGAAAGGTTACCCACACAGAGCTGACACATTAACGCAACAAATGCGCAA 3840
Db 3781 ACCGGCTGGAAAGGTTACCCACACAGAGCTGACACATTAACGCAACAAATGCGCAA 3840
QY 3841 GCAAGTATTACAAACCAACAGGTGATATCAGCGGTACGATTTCCGGGTAAACGCGTAAGT 3900
Db 3841 GCAAGTATTACAAACCAACAGGTGATATCAGCGGTACGATTTCCGGGTAAACGCGTAAGT 3900
QY 3901 GTTACGCGACTGCTGATTTAAGCACTTAATTCGGCTCAAAATTTGAAGCGAAATCGGGT 3960
Db 3901 GTTACGCGACTGCTGATTTAAGCACTTAATTCGGCTCAAAATTTGAAGCGAAATCGGGT 3960
QY 3961 GAGGCTAATGTACAAAGTCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATAGCGTA 4020
Db 3961 GAGGCTAATGTACAAAGTCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATAGCGTA 4020
QY 4021 AATGTTACGCGCAACGCTGGGATTTAAGCACTTGAAGTGGCAATGGCGGAGAAATTAATCGGACA 4080
Db 4021 AATGTTACGCGCAACGCTGGGATTTAAGCACTTGAAGTGGCAATGGCGGAGAAATTAATCGGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTCTTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTCTTAGC 4140
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTGGCTCAGATGCTAGCATCGCAGGAAGC 4200
Db 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTGGCTCAGATGCTAGCATCGCAGGAAGC 4200
QY 4201 ATTATGCTGTAATGTGACATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Db 4201 ATTATGCTGTAATGTGACATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
QY 4261 GATATTAAGCAACCAACGCGCACCTTGTTTATTAAAGCAAGATGCTAAGCTAAATGGT 4320
Db 4261 GATATTAAGCAACCAACGCGCACCTTGTTTATTAAAGCAAGATGCTAAGCTAAATGGT 4320
QY 4321 GATCATCAGGTGATGATACAGAGTGAATGAGTCAACGCAACGCGCTCTGATGTTG 4380
Db 4321 GATCATCAGGTGATGATACAGAGTGAATGAGTCAACGCAACGCGCTCTGATGTTG 4380
QY 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATGGGTTA 4440
Db 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATGGGTTA 4440
QY 4441 AATATCATTTCAAGAGTGTAGAAACACTGTGCGCTTAAGAGCGCAAGGAATTTAGGTTG 4500
Db 4441 AATATCATTTCAAGAGTGTAGAAACACTGTGCGCTTAAGAGCGCAAGGAATTTAGGTTG 4500
QY 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAGAGTGAATTTGAAGCGCAACGCGTCTT 4560
Db 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAGAGTGAATTTGAAGCGCAACGCGTCTT 4560
QY 4561 GAAAAAGTAAAAAGATTTATCTGATGAAGAGAGAGAAACATTAAGTCTTAAGTGTGTAAGT 4620
Db 4561 GAAAAAGTAAAAAGATTTATCTGATGAAGAGAGAGAAACATTAAGTCTTAAGTGTGTAAGT 4620
QY 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTAAGTCAATACAAATTAAGTCAATTAAGT 4680
Db 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTAAGTCAATACAAATTAAGTCAATTAAGT 4680
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTCAAGGTGAAGCGGTCTTCAAGTGTGTAAT 4740
Db 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTCAAGGTGAAGCGGTCTTCAAGTGTGTAAT 4740
QY 4741 GCGCACGAGTATGACCAATGTCTGACGATGGACGCGGTAGTCAAGTAATTTGACAAG 4800
Db 4741 GCGCACGAGTATGACCAATGTCTGACGATGGACGCGGTAGTCAAGTAATTTGACAAG 4800
QY 4801 GTAGATTTTCCTCTGCAATGAAGTCAATTTTCTGATTTTCTGATTTTCTGATTTTCTGATTTT 4860
Db 4801 GTAGATTTTCCTCTGCAATGAAGTCAATTTTCTGATTTTCTGATTTTCTGATTTTCTGATTTT

Db 4801 GTAGATTTTCATCTCTGCAATGAAGTCAATTTTCTGATTTTCTGATTTTCTGATTTTCTGATTTT 4860
QY 4861 GTTACAGTACGGGCTTTACCCATCTTTGTAATAAAATTAACGAGAAATTAACGAGAAATTAACGAGAAAT 4920
Db 4861 GTTACAGTACGGGCTTTACCCATCTTTGTAATAAAATTAACGAGAAATTAACGAGAAATTAACGAGAAAT 4920
QY 4921 AACAGGTTATTATTATG 4937
Db 4921 AACAGGTTATTATTATG 4937
RESULT 9
US-08-038-682-6
; Sequence 6, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-6
Query Match 97.6%; Score 4820.2; DB 1; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 1 TAAATATACAGATAATAAATAAATCAAGATTTTGTGATGACAAACAACAAATTAACA 60
Db 441 TAAATATACAGATAATAAATAAATAAATCAAGATTTTGTGATGACAAACAACAAATTAACA 500
QY 61 CACCTTTTTCGAGCTCTATATCAATAATTTTAAATAAATAGTATAAATCGGCATATAA 120
Db 501 CACCTTTTTCGAGCTCTATATCAATAATTTTAAATAAATAGTATAAATCGGCATATAA 560
QY 121 AATGATATATCTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 180
Db 561 AATGATATATCTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCAT 620
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 680

Db 2841 AGCAATAGCAAGGCTTAAACACACAGTATAGAGCTCTGAGGGTGAATTTAAACGC 2900
|||||
QY 2461 GTAATGGCAACATGTCATCAATCTCAAGAGAGGAGGAAAGTTAATTTCAATTAATA 2520
|||||
Db 2901 GTAATGGCAACATGTCATCAATCTCAAGAGAGGAGGAAAGTTAATTTCAATTAATA 2960
|||||
QY 2521 CCAACAGAGAAATGAACACAGCAACACCTTTACCAATTCGGTTTTAGCCAAATACACA 2580
|||||
Db 2961 CCAACAGAGAAATGAACACAGCAACACCTTTACCAATTCGGTTTTAGCCAAATACACA 3020
|||||
QY 2581 GCCACTGGTGGGGCTCTGTTTTTGTATATATATGCAACCAATCTTGGCAGAGGGCT 2640
|||||
Db 3021 GCCACTGGTGGGGCTCTGTTTTTGTATATATATGCAACCAATCTTGGCAGAGGGCT 3080
|||||
QY 2641 GAGTTAAAATGAGTGAATTAATATCTTAACGGCGCTAATTTTACCTTAAATTCCTAT 2700
|||||
Db 3081 GAGTTAAAATGAGTGAATTAATATCTTAACGGCGCTAATTTTACCTTAAATTCCTAT 3140
|||||
QY 2701 GTTCGGCGGATGAGCGCTTTTAAATCAACAGACTTAACCAATTAATGCAACCAATTC 2760
|||||
Db 3141 GTTCGGCGGATGAGCGCTTTTAAATCAACAGACTTAACCAATTAATGCAACCAATTC 3200
|||||
QY 2761 AATTTGAGCTCAGACAGAGCAAGATGATTTTATGACGGGTACGCAATGCCATC 2820
|||||
Db 3201 AATTTGAGCTCAGACAGAGCAAGATGATTTTATGACGGGTACGCAATGCCATC 3260
|||||
QY 2821 AATTTGAGCTCAGACAGAGCAAGATGATTTTATGACGGGTACGCAATGCCATC 3280
|||||
Db 3261 AATTTGAGCTCAGACAGAGCAAGATGATTTTATGACGGGTACGCAATGCCATC 3320
|||||
QY 2881 AGCAGCAGCATTTACGGGGAATTTACTATCAGAAAGCAATGTTACGCTAGAGGC 2940
|||||
Db 3321 AGCAGCAGCATTTACGGGGAATTTACTATCAGAAAGCAATGTTACGCTAGAGGC 3380
|||||
QY 2941 AATTAAGCCCTTAATCAGAAACATTAAGGATAGATTTAATAACTTGGCAGCTTGCTC 3000
|||||
Db 3381 AATTAAGCCCTTAATCAGAAACATTAAGGATAGATTTAATAACTTGGCAGCTTGCTC 3060
|||||
QY 3001 GTTAATGGAGTTTAAAGTAACTGGCGAAATGCAGATTTAAGGCAATCTCACTATT 3060
|||||
Db 3441 GTTAATGGAGTTTAAAGTAACTGGCGAAATGCAGATTTAAGGCAATCTCACTATT 3500
|||||
QY 3061 TCAGAAAGCGCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120
|||||
Db 3501 TCAGAAAGCGCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3560
|||||
QY 3121 ACCAATATGCACTGCCGAATTAATATACCAAGAGTGGTAAACTTGGCAATGTT 3180
|||||
Db 3561 ACCAATATGCACTGCCGAATTAATATACCAAGAGTGGTAAACTTGGCAATGTT 3620
|||||
QY 3181 ACCAATATGCACTGCCGAATTAATATACCAAGAGTGGTAAACTTGGCAATGTT 3620
|||||
Db 3621 ACCAATATGCACTGCCGAATTAATATACCAAGAGTGGTAAACTTGGCAATGTT 3620
|||||
QY 3241 GCGGAGATATATCAACAAAAGAGGCTTAAATATTAACAGACATTAATGATGCT 3300
|||||
Db 3681 GCGGAGATATATCAACAAAAGAGGCTTAAATATTAACAGACATTAATGATGCT 3740
|||||
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGCAACCTTCAAGTTTCCGAT 3360
|||||
Db 3741 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGCAACCTTCAAGTTTCCGAT 3800
|||||
QY 3361 AAAATTAATATCAACAAACAGATAACCAATCAAAAAGGCTATTGATGGAGGAGCTTAGT 3420
|||||
Db 3801 AAAATTAATATCAACAAACAGATAACCAATCAAAAAGGCTATTGATGGAGGAGCTTAGT 3860
|||||
QY 3421 TCAGATGCGCAAGTAAATGCGCACTTAATTAATAACCAAGAAATTTGAATTTGACAGAA 3480
|||||
Db 3861 TCAGATGCGCAAGTAAATGCGCACTTAATTAATAACCAAGAAATTTGAATTTGACAGAA 3920
|||||
QY 3481 GACCTTAAGTTTTCAGGTTTCAATAAAGCAGAGATTACGCCAAGAGTGGTAGAGATTTA 3540
|||||

Db 3921 GACCTAAGTATTTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGATTTA 3980
|||||
QY 3541 ACTATTGGCAACAGTAATGACGGTTAAACAGCGTCCGAGGCAAAACAGTAATTTAAC 3600
|||||
Db 3981 ACTATTGGCAACAGTAATGACGGTTAAACAGCGTCCGAGGCAAAACAGTAATTTAAC 4040
|||||
QY 3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCAATGTGACATTAATAGCAAAAGTG 3660
|||||
Db 4041 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCAATGTGACATTAATAGCAAAAGTG 4100
|||||
QY 3661 AAAACATCTAGCAGCAATGCGGACGCTGAAAGCAATAGCGACACAGTACCGGCTTAATC 3720
|||||
Db 4101 AAAACATCTAGCAGCAATGCGGACGCTGAAAGCAATAGCGACACAGTACCGGCTTAATC 4160
|||||
QY 3721 ATTACTGCAAAAATAGTAGAAGTAAACAAAGATATTACTTCTCTCAACAGTAATATC 3780
|||||
Db 4161 ATTACTGCAAAAATAGTAGAAGTAAACAAAGATATTACTTCTCTCAACAGTAATATC 4220
|||||
QY 3781 ACCGCTCGGAAAAGGTTTACCACACAGCGCTGACCAATTAACGCAACAAATGGCAAA 3840
|||||
Db 4221 ACCGCTCGGAAAAGGTTTACCACACAGCGCTGACCAATTAACGCAACAAATGGCAAA 4280
|||||
QY 3841 GCAAGTATTAACCAACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACAGGTAAGT 3900
|||||
Db 4281 GCAAGTATTAACCAACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACAGGTAAGT 4340
|||||
QY 3901 GTTAGCGGACTGGTGATTTTAAACCACTAAATCCGGCTCAAAAATTTGAAGCAAAATCGGT 3960
|||||
Db 4341 GTTAGCGGACTGGTGATTTTAAACCACTAAATCCGGCTCAAAAATTTGAAGCAAAATCGGT 4400
|||||
QY 3961 GAGGCTAATGTAAACAGTCAACAGGTACAAATTTGGGGGTACAAATTTCCGGTAAACAGGTA 4020
|||||
Db 4401 GAGGCTAATGTAAACAGTCAACAGGTACAAATTTGGGGGTACAAATTTCCGGTAAACAGGTA 4460
|||||
QY 4021 AATGTTAGGCAACCGCTGGGATTTAACAGTTTGGGAATGGCGAGAAATTAATGCGACA 4080
|||||
Db 4461 AATGTTAGGCAACCGCTGGGATTTAACAGTTTGGGAATGGCGAGAAATTAATGCGACA 4520
|||||
QY 4081 GAAGAGCTCAACCTTTAACCGCAACAGGGAATTAACCTGACTACTGAAGCGGTTCTAGC 4140
|||||
Db 4521 GAAGAGCTCAACCTTTAACCGCAACAGGGAATTAACCTGACTACTGAAGCGGTTCTAGC 4580
|||||
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTGGGTCAGATTTGTCAGATGTCGAGGAAGC 4200
|||||
Db 4581 ATCACTTCAACTAAGGCTCAGGTAGACCTTGGGTCAGATTTGTCAGATGTCGAGGAAGC 4640
|||||
QY 4201 ATTAATGCTGCTTAATGTGACATTTAAATPACTACAGGCACTTTAACCCGCTGGCAGGCTCG 4260
|||||
Db 4641 ATTAATGCTGCTTAATGTGACATTTAAATPACTACAGGCACTTTAACCCGCTGGCAGGCTCG 4700
|||||
QY 4261 GATATTAAAGCAACCGCGGCACTTGGTTTATTAAACGCAAAAGATGCTAAGCTTAATGGT 4320
|||||
Db 4701 GATATTAAAGCAACCGCGGCACTTGGTTTATTAAACGCAAAAGATGCTAAGCTTAATGGT 4760
|||||
QY 4321 GATCATPAGGTGATGATGACAGAGTGAATGACGTAAACGCAAGCGGCTCTGGTAGTGTG 4380
|||||
Db 4761 GATCATPAGGTGATGATGACAGAGTGAATGACGTAAACGCAAGCGGCTCTGGTAGTGTG 4820
|||||
QY 4381 ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGGTTA 4440
|||||
Db 4821 ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGGTTA 4880
|||||
QY 4441 AATATCATTTTCAAAAGATGTTAGCAACACTGTCGCTTAAGCGCAAGGAAATTTAGGTG 4500
|||||
Db 4881 AATATCATTTTCAAAAGATGTTAGCAACACTGTCGCTTAAGCGCAAGGAAATTTAGGTG 4940
|||||
QY 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGCAAGTAAATTAAGCGCAAGGAAATTTAGGTG 4560
|||||
Db 4941 AATATATCCAGCCAGGTGTAGCAAGTGTAGCAAGTAAATTAAGCGCAAGGAAATTTAGGTG 5000
|||||
QY 4561 GAAAAAGTAAAAAGATTTTATCTGATGAAGAAAGAAACATTAAGCTAAACTTTGGTGAAGT 4620
|||||
Db 5001 GAAAAAGTAAAAAGATTTTATCTGATGAAGAAAGAAACATTAAGCTAAACTTTGGTGAAGT 5060
|||||

QY 4621 GCTGTACGTTTGTGTTGAGCCAAATATACAAATTACAGTCAATACACAAAATGAATTTACA 4680
 Db 5061 GCTGTACGTTTGTGTTGAGCCAAATATACAAATTACAGTCAATACACAAAATGAATTTACA 5120
 QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGGCGTGTCTCAAGTGGTAAT 4740
 Db 5121 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGGCGTGTCTCAAGTGGTAAT 5180
 QY 4741 GCGCAGCAGTATGTACCAATGTTCTGACGATGACAGCGCTAGTCAAGTAAATGACAAG 4800
 Db 5181 GCGCAGCAGTATGTACCAATGTTCTGACGATGACAGCGCTAGTCAAGTAAATGACAAG 5240
 QY 4801 GTAGATTTCATCTGCAATGAAGTCAATTTATTTTCGTTATTTATTTTCGTTGTTGTTAAA 4860
 Db 5241 GTAGATTTCATCTGCAATGAAGTCAATTTATTTTCGTTATTTATTTTCGTTGTTGTTAAA 5300
 QY 4861 GTTCAGTACGGCTTTACCCATCTCTGTAAGAAATACGAGAAATACAAATAAGTATTTT 4920
 Db 5301 GTTCAGTACGGCTTTACCCATCTCTGTAAGAAATACGAGAAATACAAATAAGTATTTT 5360
 QY 4921 AACAGGTTATTATATG 4937
 Db 5361 AACAGGTTATTATATG 5377

RESULT 10

US-08-302-832-6
 : Sequence 6, Application US/08302832
 : Patent No. 5603938
 : GENERAL INFORMATION:
 : APPLICANT: Barenkamp, Stephen J
 : TITLE OF INVENTION: High Molecular Weight Surface Proteins
 : TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Shoemaker and Mattare, Ltd.
 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 : CITY: Arlington
 : STATE: Virginia
 : COUNTRY: U.S.A.
 : ZIP: 22202-0286
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/302,832
 : FILING DATE: 16-SEP-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: GB 9205704.1
 : FILING DATE: 16-MAR-1992
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US pct/us93/02166
 : FILING DATE: 16-MAR-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Berkstreser, Jerry W
 : REGISTRATION NUMBER: 22,651
 : REFERENCE/DOCKET NUMBER: 1038-404
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 415-0810
 : TELEFAX: (703) 415-0813
 : INFORMATION FOR SEQ ID NO: 6:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 9323 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : US-08-302-832-6

Query Match 97.6%; Score 4820.2; DB 1; Length 9323;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 60
 Db 441 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 500
 QY 61 CACCTTTTTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAATCGCCATATAA 120
 Db 501 CACCTTTTTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAATCGCCATATAA 560
 QY 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
 Db 561 AATGGTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620
 QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
 Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680
 QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
 Db 681 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
 QY 301 GAACGCAATGATAAAGTAATTTTAAATTTTCAACTTAACCTTTAGGAGAGAAATATGAACA 360
 Db 741 GAACGCAATGATAAAGTAATTTTAAATTTTCAACTTAACCTTTAGGAGAGAAATATGAACA 800
 QY 361 ATATATCGTCTCAAAATTCAGCAACGCTGATGCTTTGGTGTGCTGTCTGCTGAATGGCA 420
 Db 801 ATATATCGTCTCAAAATTCAGCAACGCTGATGCTTTGGTGTGCTGTCTGCTGAATGGCA 860
 QY 421 CGGGGTGTGACCATTCACAGAAAAGCCTCCCTATGTTACTATCTTTAGGTGTTAAAC 480
 Db 861 CGGGGTGTGACCATTCACAGAAAAGCCTCCCTATGTTACTATCTTTAGGTGTTAAAC 920
 QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTTAAACCTATTTTAC 540
 Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTTAAACCTATTTTAC 980
 QY 541 CAATCTGTTTACAGAGCGCTTTACAAGGAATGGATGTAGTACACGGCACACGCCACTATG 600
 Db 981 CAATCTGTTTACAGAGCGCTTTACAAGGAATGGATGTAGTACACGGCACACGCCACTATG 1040
 QY 601 CAAGTAGATGGTAATAAATAAACCCTTTATCCGCAACAGTGTGAGCGCTATCATTAATTTGAAA 660
 Db 1041 AAGAAAACAAGTAATAAATAAACCCTTTATCCGCAACAGTGTGAGCGCTATCATTAATTTGAAA 1100
 QY 661 CAATTTAACATCGACCAAAATGAATGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 720
 Db 1101 CAATTTAACATCGACCAAAATGAATGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1160
 QY 721 GTATTCACCGTGTACATCTACCAAAATCTCCCAATTTAAAGGGGATTTTAGATTTTAAAC 780
 Db 1161 GTATTCACCGTGTACATCTACCAAAATCTCCCAATTTAAAGGGGATTTTAGATTTTAAAC 1220
 QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATAGGTAAAGCAATTTATTAAC 840
 Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATAGGTAAAGCAATTTATTAAC 1280
 QY 841 ACTAATGGCTTTTACGGCTTCTACGGCTAGACATTTCTAACGAAACATCAAGCGCGCTAAT 900
 Db 1281 ACTAATGGCTTTTACGGCTTCTACGGCTAGACATTTCTAACGAAACATCAAGCGCGCTAAT 1340
 QY 901 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATCAAGGTTAAT 960
 Db 1341 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATCAAGGTTAAT 1400
 QY 961 ACTGTCGTAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGGAGGAGGAGGAGG 1020
 Db 1401 ACTGTCGTAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGGAGGAGGAGGAGG 1460

Qy	4321	GATG	CATCAGTGATGATGATACAG	AGTGAATCGAGCTCAACCGCCTCTGCTAGTGTG	4380
Db	4761	GATG	CATCAGTGATGATGATACAG	AGTGAATCGAGCTCAACCGCCTCTGCTAGTGTG	4820
Qy	4381	ACTG	GGCAACCTCAAGCAGTGTGA	ATATCACTCTGGGATTTAAACACAGATAAATGGGTTA	4440
Db	4821	ACTG	GGCAACCTCAAGCAGTGTGA	ATATCACTGGGATTTAAACACAGATAAATGGGTTA	4880
Qy	4441	AATAT	CATTTCTGAAAGATGGTAGA	AACACTCTGGCGCTTTAAGAGCAAGGAAATTTGAGGTG	4500
Db	4881	AATAT	CATTTCTGAAAGATGGTAGA	AACACTCTGGCGCTTTAAGAGCAAGGAAATTTGAGGTG	4940
Qy	4501	AAATAT	ATCCAGCCAGGTGTAGCAAG	TGTAGAAAGTAATTTGAAGCGAAACGCGTCTTT	4560
Db	4941	AAATAT	ATCCAGCCAGGTGTAGCAAG	TGTAGAAAGTAATTTGAAGCGAAACGCGTCTTT	5000
Qy	4561	GAAAAG	TAAAAGATTTATCTGATCA	AGAAAGAGAGAAACATTTAGCTAAACTTTGGTGTAAGT	4620
Db	5001	GAAAAG	TAAAAGATTTATCTGATCA	AGAAAGAGAGAAACATTTAGCTAAACTTTGGTGTAAGT	5060
Qy	4621	GCTGT	ACGTTTTGTTGAGCCAAAT	AATACAATTACAGTCAATACACAAATGAATTTTACA	4680
Db	5061	GCTGT	ACGTTTTGTTGAGCCAAAT	AATACAATTACAGTCAATACACAAATGAATTTTACA	5120
Qy	4681	ACCAG	ACCGTCAAGTCAAGTGATA	TTTCTGAAGGTAAGCGGTGTTTCTCAAGTGGTAAAT	4740
Db	5121	ACCAG	ACCGTCAAGTCAAGTGATA	TTTCTGAAGGTAAGCGGTGTTTCTCAAGTGGTAAAT	5180
Qy	4741	GGCG	CACGAGTATGTACCAATCTG	CTGACGATGACAGCCGTAGTCAAGTAATTTGACAAG	4800
Db	5181	GGCG	CACGAGTATGTACCAATCTG	CTGACGATGACAGCCGTAGTCAAGTAATTTGACAAG	5240
Qy	4801	GTAG	ATTTTCATCTCGAATGAAGT	CACTATTTTATTTTCGTATTTATTTTACHTGCTGGGTTTAAA	4860
Db	5241	GTAG	ATTTTCATCTCGAATGAAGT	CACTATTTTATTTTCGTATTTATTTTACHTGCTGGGTTTAAA	5300
Qy	4861	GTTCA	GTCAGTCGGGCTTTACCCAT	CTTTGTGTAATAAATAATTTACGGAGAAATACAATAAGTATTTTT	4920
Db	5301	GTTCA	GTCAGTCGGGCTTTACCCAT	CTTTGTGTAATAAATAATTTACGGAGAAATACAATAAGTATTTTT	5360
Qy	4921	AACAG	GTATTATTATTATG	4937	
Db	5361	AACAG	GTATTATTATTATG	5377	

RESULT 11
US-08-530-198-6

```

US-08-530-198-6
; sequence 6, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-6

Query Match      97.6%; Score 4820.2; DB 2; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGATTAATAAATAATCAAGATTTTGTGATGACAAACAAATTCACAA 60
DB 441 TAAATATACAGATTAATAAATAATCAAGATTTTGTGATGACAAACAAATTCACAA 500
QY 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATA 120
DB 501 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATA 560
QY 121 AATGATATATCTTTTCATCTTTTCAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
DB 561 AATGATATATCTTTTCATCTTTTCAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620
QY 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
DB 621 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 680
QY 241 CACATGAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
DB 681 CACATGAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
QY 301 GACGCAATGATGAATGAATTTAATTTGTTCACTAACCTTTAGGAGGAGGAGGAGGAGGAGG 360
DB 741 GACGCAATGATGAATGAATTTAATTTGTTCACTAACCTTTAGGAGGAGGAGGAGGAGGAGG 800
QY 361 ATATATCGTCTCAAAATTCAGCAACGCTGAATCTTTGGTGTCTGTCTGAATTTGCA 420
DB 801 ATATATCGTCTCAAAATTCAGCAACGCTGAATCTTTGGTGTCTGTCTGAATTTGCA 860
QY 421 CCGGGTTGTGACCATTCACAGAAAGGCTTCGCTGATGTTACTATCTTTAGTGTGAAC 480
DB 861 CCGGGTTGTGACCATTCACAGAAAGGCTTCGCTGATGTTACTATCTTTAGTGTGAAC 920
QY 481 CACTTAGCGTTAAAGCCACCTTTCCGCTATGTTACTATCTTTAGTGTGAACATCTATTCCA 540
DB 921 CACTTAGCGTTAAAGCCACCTTTCCGCTATGTTACTATCTTTAGTGTGAACATCTATTCCA 980
QY 541 CAATCTGTTTACGACGGCTTTACAGGAATGATGATGATACACGCGACAGCCCATG 600
DB 981 CAATCTGTTTACGACGGCTTTACAGGAATGATGATGATGATGATGATGATGATGATGATGAT 1040
QY 601 CAAGTAGTGGTAAATAAACCCTTTATCCGCAACAGTGTGACGCTATCATTAATTCGAAA 660
DB 1041 AAGAAAACAAGTAATAAACCATTTATCCGCAACAGTGTGACGCTATCATTAATTCGAAA 1100
QY 661 CAATTTACATCGACCAAAATGAATGGTGGAGTTTTCAGAAAACAACTCCGCC 720
DB 1101 CAATTTACATCGACCAAAATGAATGGTGGAGTTTTCAGAAAACAACTCCGCC 1160
QY 721 GTATTCACCGGTGTTACATCTAACCAATCTCCCAATTTAAAGGGATTTAGATTTCTAAC 780
DB 1161 GTATTCACCGGTGTTACATCTAACCAATCTCCCAATTTAAAGGGATTTAGATTTCTAAC 1220
QY 781 GGNCAAGTCTTTTATCAACCAAAAGGATGATCAATAGGTAAGAGCGCAATTAATTAAC 840

```

```

|||||
1221 GGCAAGTCTTTTAAATCAACCCAAATGGTATCACAATAGTAAAGACCAATTAATTAAC 1280
QY 841 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 900
DB 1281 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 1340
QY 901 TTACCTTTCAGCAACCAACAGATTAAGCGCTCGCTGAAATTTGTGAATCAGCGTTAAAT 960
DB 1341 TTACCTTTCAGCAACCAACAGATTAAGCGCTCGCTGAAATTTGTGAATCAGCGTTAAAT 1400
QY 961 ACTGTCGTTAAAGACGCGCAGTGTAAATCTTATTGTTGGCAAGTGAAGAACGAGGGTGTG 1020
DB 1401 ACTGTCGTTAAAGACGCGCAGTGTAAATCTTATTGGTGGCAAGTGAAGAACGAGGGTGTG 1460
QY 1021 ATTAGCGTAATTTGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAATCACCATCAGCGAT 1080
DB 1461 ATTAGCGTAATTTGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAATCACCATCAGCGAT 1520
QY 1081 ATAATAAACCCCAACCACTTACTTACAGCATTTGCCGCTGAAATGAAGCGGTCAATCTG 1140
DB 1521 ATAATAAACCCCAACCACTTACTTACAGCATTTGCCGCTGAAATGAAGCGGTCAATCTG 1580
QY 1141 GCGCATATTTTTCGCAAGCGGTACATTAATGTCGCTGCCACTATTCGAAACCAA 1200
DB 1581 GCGCATATTTTTCGCAAGCGGTACATTAATGTCGCTGCCACTATTCGAAACCAA 1640
QY 1201 GSTAAACTTTCGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTGTTCTTCCGCC 1260
DB 1641 GSTAAACTTTCGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTGTTCTTCCGCC 1700
QY 1261 AAGAGGTTGAAGCGGAAATTTGCCGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
DB 1701 AAGAGGTTGAAGCGGAAATTTGCCGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1760
QY 1321 GCGAGCTGATGATTACAGCGCATAAAGTCAATTAATAACAGGTGCGAGTTATCGACCTT 1380
DB 1761 GCGAGCTGATGATTACAGCGCATAAAGTCAATTAATAACAGGTGCGAGTTATCGACCTT 1820
QY 1381 TCAGTGAAGAGGCGGAGAACTTACCTTGGCGGTGAGAGCGCGGAGGAGGTAAAGAC 1440
DB 1821 TCAGTGAAGAGGCGGAGAACTTACCTTGGCGGTGAGAGCGCGGAGGAGGTAAAGAC 1880
QY 1441 GCGATTCAATTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500
DB 1881 GCGATTCAATTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1940
QY 1501 AAGAAAAGCGCGACGCGCTATTGTTGGGGCGATATTGCGTTAATTGACGGCAATATT 1560
DB 1941 AAGAAAAGCGCGACGCGCTATTGTTGGGGCGATATTGCGTTAATTGACGGCAATATT 2000
QY 1561 AACGCTCAAGTGTAGTGTGATATCCGTAACCGGTGGTTTTTGGAGACATCGGGGCAT 1620
DB 2001 AACGCTCAAGTGTAGTGTGATATCCGTAACCGGTGGTTTTTGGAGACATCGGGGCAT 2060
QY 1621 TATTATCCATTCACAGCAATGCAATTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
DB 2061 TATTATCCATTCACAGCAATGCAATTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2120
QY 1681 GATGTACAAATTTAGCGCGAAGACCCCTTCGCAATTAATACCGGTATATAATGATGAATTC 1740
DB 2121 GATGTACAAATTTAGCGCGAAGACCCCTTCGCAATTAATACCGGTATATAATGATGAATTC 2180
QY 1741 CCAAGAGCGACCGGTGAAGCAAGGAGCCCTTAAAAAATAGGCAACTCAAAACACGCTA 1800
DB 2181 CCAAGAGCGACCGGTGAAGCAAGGAGCCCTTAAAAAATAGGCAACTCAAAACACGCTA 2240
QY 1801 ACCAATACACTATTTCATTAATTTCTGAAAACCGCTGGACAAATGAATATAAGCGCATCA 1860
DB 2241 ACCAATACACTATTTCATTAATTTCTGAAAACCGCTGGACAAATGAATATAAGCGCATCA 2300
QY 1861 AGAAACCTTACCCTTAAATAGTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920
|||||

```

Db 2301 AGAAACTTACCGTTAATAGCTCAATCAACATCGAAGCAACTCCCACTTAAATCTCCAT 2360
QY 1921 AGTAAAGGTCAGCGTGGCGAGGCGTTCAGATTGATGGAGATATCTCTCTAAAGCGGA 1980
Db 2361 AGTAAAGGTCAGCGTGGCGAGGCGTTCAGATTGATGGAGATATCTCTCTAAAGCGGA 2420
QY 1981 AATTTAACCAATTTATCTCGCGATGGGTTGATGTTTCATATAAATATATAGCGTTGATCAG 2040
Db 2421 AATTTAACCAATTTATCTCGCGATGGGTTGATGTTTCATATAAATATATAGCGTTGATCAG 2480
QY 2041 GGTGTTTTAAATATACCGCGCGTCCGTTAGCTTTTGAAGGTGGAATAACAAAGCACGC 2100
Db 2481 GGTGTTTTAAATATACCGCGCGTCCGTTAGCTTTTGAAGGTGGAATAACAAAGCACGC 2540
QY 2101 GACCGGCAAAATGCTAAAATGTCGCCAGGGCACTGTAAACCATTTACAGGAGAGGGAATA 2160
Db 2541 GACCGGCAAAATGCTAAAATGTCGCCAGGGCACTGTAAACCATTTACAGGAGAGGGAATA 2600
QY 2161 GATTTACAGGCTAACACGCTATCTTTAAACGGAAACGGGTAAAGGTCTGTAATATCATTTCA 2220
Db 2601 GATTTACAGGCTAACACGCTATCTTTAAACGGAAACGGGTAAAGGTCTGTAATATCATTTCA 2660
QY 2221 TCAGTGAATTAATTAACCCACAATCTTAGTGCCACAAATTAACATATCTGGGAATATAACA 2280
Db 2661 TCAGTGAATTAATTAACCCACAATCTTAGTGCCACAAATTAACATATCTGGGAATATAACA 2720
QY 2281 ATTAACCAAACTACGAGAAGACACTCGTATTGGCAACACGACCATGATTCGCACGTG 2340
Db 2721 ATTAACCAAACTACGAGAAGACACTCGTATTGGCAACACGACCATGATTCGCACGTG 2780
QY 2341 AAGCTCAGTCTCTTAATCTAGACAGAGGCGCAAAATTTTACCTTTATTAATACATTTCA 2400
Db 2781 AAGCTCAGTCTCTTAATCTAGACAGAGGCGCAAAATTTTACCTTTATTAATACATTTCA 2840
QY 2401 AGCAATAGCAAGGCTTACACACAGTATAGAAGTCTGCGAGGTTGAAATTTTAAACGGC 2460
Db 2841 AGCAATAGCAAGGCTTACACACAGTATAGAAGTCTGCGAGGTTGAAATTTTAAACGGC 2900
QY 2461 GTAAATGGCAACATGTCAATCAATCTCAAGAAAGAGCGGAAAGTTTAAATTTCAAATTAATA 2520
Db 2901 GTAAATGGCAACATGTCAATCAATCTCAAGAAAGAGCGGAAAGTTTAAATTTCAAATTAATA 2960
QY 2521 CCACAGAGAACATGATCAACACAGCAACCTTTTACCAATTCGGTTTATAGCCAAATATCACA 2580
Db 2961 CCACAGAGAACATGATCAACACAGCAACCTTTTACCAATTCGGTTTATAGCCAAATATCACA 3020
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGGCAGAGGGCT 2640
Db 3021 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGGCAGAGGGCT 3080
QY 2641 GAGTTAAATAGTCAAAATTAATATCTTAAGCGGCTAAATTTTACCTTAAATTCCTCAT 2700
Db 3081 GAGTTAAATAGTCAAAATTAATATCTTAAGCGGCTAAATTTTACCTTAAATTCCTCAT 3140
QY 2701 GTTCGGGCGATGACGCTTTTAAATCAACAAGACTTAAACATAAATGCAACCAATTTCA 2760
Db 3141 GTTCGGGCGATGACGCTTTTAAATCAACAAGACTTAAACATAAATGCAACCAATTTCA 3200
QY 2761 AATTTACGCTCAGACAGACGAAGATGATTTTATGACGGGTACGACGCAATGCGCATC 2820
Db 3201 AATTTACGCTCAGACAGACGAAGATGATTTTATGACGGGTACGACGCAATGCGCATC 3260
QY 2821 AATTTACCTTACAACATATCCATTTCTGGGCGGTAAATGTCAACCTTGGTGGACAAACTCA 2880
Db 3261 AATTTACCTTACAACATATCCATTTCTGGGCGGTAAATGTCAACCTTGGTGGACAAACTCA 3320
QY 2881 AGCAGCAGCATTTACGGGGAATTTACTATCTGAGAAAGCAGCAAACTTACGCTAGAGCC 2940
Db 3321 AGCAGCAGCATTTACGGGGAATTTACTATCTGAGAAAGCAGCAAACTTACGCTAGAGCC 3380
QY 2941 AATAAGCCCCCTAATCAGCAAAACATAGGGATAGAGTTTATAAACTTGGCAGCTTGCTC 3000
Db 3381 AATAAGCCCCCTAATCAGCAAAACATAGGGATAGAGTTTATAAACTTGGCAGCTTGCTC 3440

QY 3001 GTTAAATGGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTTAAAGCAATCTCACATAT 3060
Db 3441 GTTAAATGGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTTAAAGCAATCTCACATAT 3500
QY 3061 TCAGAAAGCCGCACTTTTAAAGAAAGACTAGAGATACCCCTAAATATACACGGCAATTTT 3120
Db 3501 TCAGAAAGCCGCACTTTTAAAGAAAGACTAGAGATACCCCTAAATATACACGGCAATTTT 3560
QY 3121 ACCAATATGGCACTGCGCAAAATTAATATACCAAGAGAGTGGTAAACCTTTGGCAATGTT 3180
Db 3561 ACCAATATGGCACTGCGCAAAATTAATATACCAAGAGAGTGGTAAACCTTTGGCAATGTT 3620
QY 3181 ACCAATGATGTTGATTTTAAACATTTACCACCTACACCTAAACGCAACCAAGAGCATCATC 3240
Db 3621 ACCAATGATGTTGATTTTAAACATTTACCACCTACACCTAAACGCAACCAAGAGCATCATC 3680
QY 3241 GCGGAGATATTAATCAACAAAGAGAGCTTAAATATTTACAGACAGTAAATATGATGCT 3300
Db 3681 GCGGAGATATTAATCAACAAAGAGAGCTTAAATATTTACAGACAGTAAATATGATGCT 3740
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGAGCTTAAATTTTCCGAT 3360
Db 3741 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGAGCTTAAATTTTCCGAT 3800
QY 3361 ABAATTAATTCACCAACAGATTAACATCAAAAGAGTATTTGATGGAGAGCTCTAGT 3420
Db 3801 ABAATTAATTCACCAACAGATTAACATCAAAAGAGTATTTGATGGAGAGCTCTAGT 3860
QY 3421 TCAGATGCGCAAGTAAATGCAACCTAACTATTTAAACCAAAAGAAATTTGAAATTTGACAGAA 3480
Db 3861 TCAGATGCGCAAGTAAATGCAACCTAACTATTTAAACCAAAAGAAATTTGAAATTTGACAGAA 3920
QY 3481 GACCTAAGTATTTTACAGGTTTCAATTAAGAGAGATTTACAGCAAAAGTGGTAGAGATTTA 3540
Db 3921 GACCTAAGTATTTTACAGGTTTCAATTAAGAGAGATTTACAGCAAAAGTGGTAGAGATTTA 3980
QY 3541 ACTATTGGCAACAGTAAATGACGGTTAACAGCGTCCGCAAGCCAAACAGTAACCTTTTAAAC 3600
Db 3981 ACTATTGGCAACAGTAAATGACGGTTAACAGCGTCCGCAAGCCAAACAGTAACCTTTTAAAC 4040
QY 3601 AATGTTAAAGATTCACAAATCTCTGCTGACGGTCAACATGTGACACTTAAATAGCAAGTG 3660
Db 4041 AATGTTAAAGATTCACAAATCTCTGCTGACGGTCAACATGTGACACTTAAATAGCAAGTG 4100
QY 3661 ABAACATCTAGCAACATGCGGACGTGAAGCAATAGCAGCAACGATACCGGCTTAACT 3720
Db 4101 ABAACATCTAGCAACATGCGGACGTGAAGCAATAGCAGCAACGATACCGGCTTAACT 4160
QY 3721 ATTACTGCAAAATATGTAGAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATC 3780
Db 4161 ATTACTGCAAAATATGTAGAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATC 4220
QY 3781 ACCGCGTCGGAAGGTTTACCAACAGCAGCGGTGACCATTTAACGCAACAAATGCGCAA 3840
Db 4221 ACCGCGTCGGAAGGTTTACCAACAGCAGCGGTGACCATTTAACGCAACAAATGCGCAA 4280
QY 3841 GCAAGTATTTACAAACCAACAGGTTGATATCAGCGGTACGATTTCCCGTTAACACGCTAAGT 3900
Db 4281 GCAAGTATTTACAAACCAACAGGTTGATATCAGCGGTACGATTTCCCGTTAACACGCTAAGT 4340
QY 3901 GTTAGCGCACTGCTGATTTTAAACCTTAAATCCGGCTCAAAATTTGAAGCAATCCGGT 3960
Db 4341 GTTAGCGCACTGCTGATTTTAAACCTTAAATCCGGCTCAAAATTTGAAGCAATCCGGT 4400
QY 3961 GAGGCTTAATGTACAAAGTCAACAGGTACAAATTTGGCGGTACAAATTTCCCGTTAATACGGTA 4020
Db 4401 GAGGCTTAATGTACAAAGTCAACAGGTACAAATTTGGCGGTACAAATTTCCCGTTAATACGGTA 4460
QY 4021 AATGTTACGCAACCGCTGCGGATTTTAAACAGTTTGGGAATGGCGAGAAATTTAATGCGACA 4080
Db 4461 AATGTTACGCAACCGCTGCGGATTTTAAACAGTTTGGGAATGGCGAGAAATTTAATGCGACA 4520

Db 861 CGGGTGTGACCATTTCCACAGAAAGGACGGGAAAACCTGCTCGCATGAAGTGGCT 920
Qy 481 CACTTAGCGTTAAAGCCACTTTCGGCTATGTTACTATCTTTTAGGTGTAACTATCTATTCGA 540
Db 921 CACTTAGCGTTAAAGCCACTTTCGGCTATGTTACTATCTTTTAGGTGTAACTATCTATTCGA 980
Qy 541 CAATCTGTTTTAGCAAGCGGCTTACAAAGGAATGGATGTAGTACAGCGGCACACCCACTATG 600
Db 981 CAATCTGTTTTAGCAAGCGGCAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTTAC 1040
Qy 601 CAAGTAGATGTTAATAAACCATTATCCGCACACAGTGTGGAGCTATCATTAATTTGGAAA 660
Db 1041 AAGAAAACAAGTAATAAACCATTATCCGCACACAGTGTGGAGCTATCATTAATTTGGAAA 1100
Qy 661 CAATTTAAACATCCACCAAAATGAAATGGTGCAGTTTTTACAGAAAACAACAACCTCCGC 720
Db 1101 CAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTTACAGAAAACAACAACCTCCGC 1160
Qy 721 GTATTCACCGCTGTACATCTAACCAAACTCCCAATTTAAAGGGATTTTAGATTTCTAAC 780
Db 1161 GTATTCACCGCTGTACATCTAACCAAACTCCCAATTTAAAGGGATTTTAGATTTCTAAC 1220
Qy 781 GGACAGCTTTTTTAATCAACCCAAATGGTATCACAAATAGGTAAAGCGCAATTTATTAAC 840
Db 1221 GGACAGCTTTTTTAATCAACCCAAATGGTATCACAAATAGGTAAAGCGCAATTTATTAAC 1280
Qy 841 ACTAATGGCTTTACGGCTCTACGCTAGACATTTCTAACGAAAACATCAAGGGCGCTAAT 900
Db 1281 ACTAATGGCTTTACGGCTCTACGCTAGACATTTCTAACGAAAACATCAAGGGCGCTAAT 1340
Qy 901 TTCACTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACGGTTTTAAT 960
Db 1341 TTCACTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACGGTTTTAAT 1400
Qy 961 ACTGTCGGTAAAGACGCGAGTAAATCTTATTGGTGGCAAAAGTGAAGAACGAGGGTGTG 1020
Db 1401 ACTGTCGGTAAAGACGCGAGTAAATCTTATTGGTGGCAAAAGTGAAGAACGAGGGTGTG 1460
Qy 1021 ATTAGCTTAATGTGCGACATTTCTTTACTGCGAGGCAAAAATCACCATCACGAT 1080
Db 1461 ATTAGCTTAATGTGCGAGCATTTCTTTACTGCGAGGCAAAAATCACCATCACGAT 1520
Qy 1081 ATATAAACCACCAATTAATTAACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140
Db 1521 ATATAAACCACCAATTAATTAACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1580
Qy 1141 GCGGATATTTTGCAGAGCGGTAACATTAATGTCGCTGCTGCCACTATTCGAAACCAA 1200
Db 1581 GCGGATATTTTGCAGAGCGGTAACATTAATGTCGCTGCTGCCACTATTCGAAACCAA 1640
Qy 1201 GGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAAGCGGCAATTTGTTCTTTCCGCC 1260
Db 1641 GGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAAGCGGCAATTTGTTCTTTCCGCC 1700
Qy 1261 AAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1320
Db 1701 AAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1760
Qy 1321 GGCAAGCTGATGATTACAGGCGATAAAGTACATTAATAAAGAGTGCAGTTATCCAGCTT 1380
Db 1761 GGCAAGCTGATGATTACAGGCGATAAAGTACATTAATAAAGAGTGCAGTTATCCAGCTT 1820
Qy 1381 TCAGTTAAAGAGGGGAAACCTTACCTTGGCGGTGACAGCGGCGGGAAGGTAAAGAAC 1440
Db 1821 TCAGTTAAAGAGGGGAAACCTTACCTTGGCGGTGACAGCGGCGGGAAGGTAAAGAAC 1880
Qy 1441 GCATTCATTAAGCAAAACCACTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500
Db 1881 GCATTCATTAAGCAAAACCACTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1940
Qy 1501 AAAGAAAAGCGGCGGCTATTTGTTGGGCGATATTTGGGTTAAATGACGGCAATATT 1560
Db 1941 AAAGAAAAGCGGCGGCTATTTGTTGGGCGATATTTGGGTTAAATGACGGCAATATT 2000

Qy 1561 AACGCTCAAGCTAGTGTGATATCGCTAAACACCGTGGTTTTTGTGGAGACATCGGGCAT 1620
Db 2001 AACGCTCAAGCTAGTGTGATATCGCTAAACACCGTGGTTTTTGTGGAGACATCGGGCAT 2060
Qy 1621 TATTTATCCATTCAGACGAATGCTTTAAACAAAGAGTGTGCTAGACCTGAT 1680
Db 2061 TATTTATCCATTCAGACGAATGCTTTAAACAAAGAGTGTGCTAGACCTGAT 2120
Qy 1681 GATGTAACAATTTGAAGCGGAAGACCCCTTCGCAATATATACCGGTATTAATGATGATTC 1740
Db 2121 GATGTAACAATTTGAAGCGGAAGACCCCTTCGCAATATATACCGGTATTAATGATGATTC 2180
Qy 1741 CCACAGGACCGGTGAAGCAAGCGACCTTAAAAAATAGCGAACTCAAAACACGGTA 1800
Db 2181 CCACAGGACCGGTGAAGCAAGCGACCTTAAAAAATAGCGAACTCAAAACACGGTA 2240
Qy 1801 ACCAATACAACCTATTTCAAAATTTATCTGAAAACCGCTGGACAATGAATATAACGCAATCA 1860
Db 2241 ACCAATACAACCTATTTCAAAATTTATCTGAAAACCGCTGGACAATGAATATAACGCAATCA 2300
Qy 1861 AGAAAACCTTACCGTTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCCAT 1920
Db 2301 AGAAAACCTTACCGTTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCCAT 2360
Qy 1921 AGTAAAGTCAAGCTGGCGGAGCGTTCAGATTGATGGAGATATTACTTTAAAGCGGA 1980
Db 2361 AGTAAAGTCAAGCTGGCGGAGCGTTCAGATTGATGGAGATATTACTTTAAAGCGGA 2420
Qy 1981 AATTTAAACATTTATTCGGCGGATGGTGTGATTTTCATAAAAATATTACGCTTCATCAG 2040
Db 2421 AATTTAAACATTTATTCGGCGGATGGTGTGATTTTCATAAAAATATTACGCTTCATCAG 2480
Qy 2041 GGTTTTTTAAATATTACCGCCCTTCCGCTAGCTTTTGAAGTGGAAATTAACAAACGACGC 2100
Db 2481 GGTTTTTTAAATATTACCGCCCTTCCGCTAGCTTTTGAAGTGGAAATTAACAAACGACGC 2540
Qy 2101 GACGCGGCAAAATGCTAAAAATTTGCGCCAGGCACTGTAAACCATTTACAGGAGAGGAAAA 2160
Db 2541 GACGCGGCAAAATGCTAAAAATTTGCGCCAGGCACTGTAAACCATTTACAGGAGAGGAAAA 2600
Qy 2161 GATTTTCAGGCTTAACAAACGATATCTTTAAACGGAACGGGTAAAAGGCTGAATATCATTTCA 2220
Db 2601 GATTTTCAGGCTTAACAAACGATATCTTTAAACGGAACGGGTAAAAGGCTGAATATCATTTCA 2660
Qy 2221 TCAGTCAATAATTTAAACCCACAACTTTAGTGGCACAATTTAAACATATCTGGGAATATAACA 2280
Db 2661 TCAGTCAATAATTTAAACCCACAACTTTAGTGGCACAATTTAAACATATCTGGGAATATAACA 2720
Qy 2281 ATTTAAACCAACTACGAGAAGACACCTCGTATTGGCAACCGCATGATTCGCACTGG 2340
Db 2721 ATTTAAACCAACTACGAGAAGACACCTCGTATTGGCAACCGCATGATTCGCACTGG 2780
Qy 2341 AACGTCAGTCTCTTAATCTAGAGACGGCGCAATTTTACCTTTTAAATACATTTCA 2400
Db 2781 AACGTCAGTCTCTTAATCTAGAGACGGCGCAATTTTACCTTTTAAATACATTTCA 2840
Qy 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAAGTCTGACGGGTGAATTTTACAGGC 2460
Db 2841 AGCAATAGCAAGGCTTAAACACACAGTATAGAAAGTCTGACGGGTGAATTTTACAGGC 2900
Qy 2461 GTAAATGGCAACATGTCTCAATCTCAAAAGAGGCGGAAAGTTTAAATTTCAATTTAAAA 2520
Db 2901 GTAAATGGCAACATGTCTCAATCTCAAAAGAGGCGGAAAGTTTAAATTTCAATTTAAAA 2960
Qy 2521 CCAACGAGAACATGAACCAACCAACCTTTACCAATTCGGTTTTTGTAGCCAAATATCACA 2580
Db 2961 CCAACGAGAACATGAACCAACCAACCTTTACCAATTCGGTTTTTGTAGCCAAATATCACA 3020
Qy 2581 GCCACTGGTGGGCTCTGTTTTTTTGTATATATGCCAACCATTTCTGCGCAGGGGCT 2640
Db 3021 GCCACTGGTGGGCTCTGTTTTTTTGTATATATGCCAACCATTTCTGCGCAGGGGCT 3080

Db 5241 GTAGATTTTCATCCTGCAATGAAGTCATTTTATTTTCGTATTTTACTGTGGTTAA 5300
 Qy 4861 GTTCAGTAGGGCTTTACCATCTTGTAAAAAATACGAGAAATACAAATAAGTATTTT 4920
 Db 5301 GTTCAGTAGGGCTTTACCATCTTGTAAAAAATACGAGAAATACAAATAAGTATTTT 5360
 Qy 4921 AACAGGTTATTTATG 4937
 Db 5361 AACAGGTTATTTATG 5377

RESULT 13
 US-08-728-470-6
 ; Sequence 6, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728.470
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302.832
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstesser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-633
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9323 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-728-470-6

Query Match
 Best Local Similarity 97.6%; Score 4820.2; DB 2; Length 9323;
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 1 TAAATATACAAAGTAAATAAATAAATCAAGATTTTGTGATGACAAAACAAATACAA 60
 Db 441 TAAATATACAAAGTAAATAAATAAATAAATCAAGATTTTGTGATGACAAAACAAATACAA 500
 Qy 61 CACCTTTTTCGAGCTATATGCAATATTTTAAAAAATAAGTATATAATCCGCCATATA 120
 Db 501 CACCTTTTTCGAGCTATATGCAATATTTTAAAAAATAAGTATATAATCCGCCATATA 1640

Qy 121 AATGGTATATCTTTTCATCTTTTCATCTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
 Db 561 AATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620
 Qy 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
 Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680
 Qy 241 CACATGAATATGAAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
 Db 681 CACATGAATATGAAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
 Qy 301 GAACGCAATGATAAAGTAAATTTAAATTTGTTCAACTAACTTAGGAGGAGGAGGAGGAGGAGGAGGAG 360
 Db 741 GAACGCAATGATAAAGTAAATTTAAATTTGTTCAACTAACTTAGGAGGAGGAGGAGGAGGAGGAGGAG 800
 Qy 361 ATATATCGTCTCAAAATTCAGCAAAACGCTTGAATGCTTTGGTGTCTGCTGCTGCTGCTGCTGCTGCTG 420
 Db 801 ATATATCGTCTCAAAATTCAGCAAAACGCTTGAATGCTTTGGTGTCTGCTGCTGCTGCTGCTGCTGCTG 860
 Qy 421 CGGGTGTGACCATTCACAGAAAAGGCTTCCGCTATGTTACTATCTTTAGTGCTGCTGCTGCTGCTGCTG 480
 Db 861 CGGGTGTGACCATTCACAGAAAAGGCTTCCGCTATGTTACTATCTTTAGTGCTGCTGCTGCTGCTGCTG 920
 Qy 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGCTGCTGCTGCTGCTGCTGCTGCTG 980
 Qy 541 CAATCTGTTTACGAAAGCGCTTACAAAGAAATGGATGTAGTACACGGCACACGACCTATG 600
 Db 981 CAATCTGTTTACGAAAGCGCAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTTTTAC 1040
 Qy 601 CAAGTAGATGTTAAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGTTGAA 660
 Db 1041 AAGAAAACAAGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGTTGAA 1100
 Qy 661 CAATTTAATCATCGACCAAAATGAAATGGTGCAGTTTTTACAAAGAAAACAAACAACTCCGCC 720
 Db 1101 CAATTTAATCATCGACCAAAATGAAATGGTGCAGTTTTTACAAAGAAAACAAACAACTCCGCC 1160
 Qy 721 GTATTTCAACCGTGTATCATCTAACCAAAATCTCCCAATTTAAAGGGATTTTATAGTCTTAAC 780
 Db 1161 GTATTTCAACCGTGTATCATCTAACCAAAATCTCCCAATTTAAAGGGATTTTATAGTCTTAAC 1220
 Qy 781 GGACAAAGTCTTTTAAATCAACCAAAATGGTATCAGATAGTAAAGCAATATTAAC 840
 Db 1221 GGACAAAGTCTTTTAAATCAACCAAAATGGTATCAGATAGTAAAGCAATATTAAC 1280
 Qy 841 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAACATCAAGGGCGGTAAT 900
 Db 1281 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAACATCAAGGGCGGTAAT 1340
 Qy 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCAAGGTTTAAAT 960
 Db 1341 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCAAGGTTTAAAT 1400
 Qy 961 ACTGTCCGTTAAGACGCGCTGTAAATCTTTATTTGGTGGCAAGTGAAGGAGGAGGAGGAGGAGGAGG 1020
 Db 1401 ACTGTCCGTTAAGACGCGCTGTAAATCTTTATTTGGTGGCAAGTGAAGGAGGAGGAGGAGGAGGAGG 1460
 Qy 1021 ATTTAGGTAATAGTGGTGGAGCATTTCTTTACTCGCGGGGCAAAATACCATCAGCAT 1080
 Db 1461 ATTTAGGTAATAGTGGTGGAGCATTTCTTTACTCGCGGGGCAAAATACCATCAGCAT 1520
 Qy 1081 ATAAATAAACCCCAACCATTTACTTACAGCATTTCCGCGGCTGAAATGAAGGGGCTCAATCTG 1140
 Db 1521 ATAAATAAACCCCAACCATTTACTTACAGCATTTCCGCGGCTGAAATGAAGGGGCTCAATCTG 1580
 Qy 1141 GCGGATATTTTTCGCAAAAGCGGTAACATTAATTTGTCGCTGCTGCCACTATTTTCGAAACCAA 1200
 Db 1581 GCGGATATTTTTCGCAAAAGCGGTAACATTAATTTGTCGCTGCTGCCACTATTTTCGAAACCAA 1640

us-10-092-880-3.rni

Tue Mar 25 08:39:01 2003

Db 3801 AAATTAATATACCAACACAGATAAACAATCAAAAAGGGTATTGATGGAGAGGACTCTAGT 3860
 Qy 3421 TCAGATCGCAGCAAGTAAATGCCAACCTAACTATTAAACACCAAGAAATGAATTTGACAGAA 3480
 Db 3861 TCAGATCGCAGCAAGTAAATGCCAACCTAACTATTAAACACCAAGAAATGAATTTGACAGAA 3920
 Qy 3481 GACCTAAGTATTTAGGTTTCAATTAAGACAGAGATTACAGCAAAAGATGCTAGAGATTTA 3540
 Db 3921 GACCTAAGTATTTAGGTTTCAATTAAGACAGAGATTACAGCAAAAGATGCTAGAGATTTA 3980
 Qy 3541 ACTATTGGCAACAGTAATGACGGTTACAGCGTGCAGCGTGCAGCAAAAGATGCTTTTAAAC 3600
 Db 3981 ACTATTGGCAACAGTAATGACGGTTACAGCGTGCAGCGTGCAGCAAAAGATGCTTTTAAAC 4040
 Qy 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAATGTGACACTTAATAGCAAAAGTG 3660
 Db 4041 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAATGTGACACTTAATAGCAAAAGTG 4100
 Qy 3661 AAAACATCTAGCAGCAATGGCGAGCTGAAAGCAATAGCGCAACAGTACCGGCTTAACT 3720
 Db 4101 AAAACATCTAGCAGCAATGGCGAGCTGAAAGCAATAGCGCAACAGTACCGGCTTAACT 4160
 Qy 3721 ATTACTGCAAAAATCTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAAATATC 3780
 Db 4161 ATTACTGCAAAAATCTAGAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAAATATC 4220
 Qy 3781 ACCGGTCGCAAAAGGTTACCAACAGCAGCGCTGACCAATTAAGCGCAACAAATGGCAAA 3840
 Db 4221 ACCGGTCGCAAAAGGTTACCAACAGCAGCGCTGACCAATTAAGCGCAACAAATGGCAAA 4280
 Qy 3841 GCAAGTATTACACCAACAGTATATACAGCGTACGATTTCCGGTAAACAGGTAAGT 3900
 Db 4281 GCAAGTATTACACCAACAGTATATACAGCGTACGATTTCCGGTAAACAGGTAAGT 4340
 Qy 3901 GTTAGCGGACTGGTATTAACCACTTAATCCGGCTCAAAAATTAAGCGAAATCGGGT 3960
 Db 4341 GTTAGCGGACTGGTATTAACCACTTAATCCGGCTCAAAAATTAAGCGAAATCGGGT 4400
 Qy 3961 GAGGCTAATGTAACAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATAGGTA 4020
 Db 4401 GAGGCTAATGTAACAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATAGGTA 4460
 Qy 4021 AATGTTACGCAAAACGCTGGCGATTTAAACAGTTGGGAATGGCGCAAAATTAATGCGACA 4080
 Db 4461 AATGTTACGCAAAACGCTGGCGATTTAAACAGTTGGGAATGGCGCAAAATTAATGCGACA 4520
 Qy 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGTTCTAGC 4140
 Db 4521 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGTTCTAGC 4580
 Qy 4141 ATCACTTCAACTAAGGGTCAAGTACCTTTGGCTCAGAAATGTAAGTACGAGCAAGC 4200
 Db 4581 ATCACTTCAACTAAGGGTCAAGTACCTTTGGCTCAGAAATGTAAGTACGAGCAAGC 4540
 Qy 4201 ATTAATGCTGCTAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
 Db 4641 ATTAATGCTGCTAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4700
 Qy 4261 GATATTAAAGCAACAGCGGCACCTTGGTTATTAAACGCAAAAGATGCTAAGCTAAATGGT 4320
 Db 4701 GATATTAAAGCAACAGCGGCACCTTGGTTATTAAACGCAAAAGATGCTAAGCTAAATGGT 4760
 Qy 4321 GATGCACTAGGTATAGTACAGAGTGAATGAGTCAAGCAAGCGGCTCTGGTAGTGTG 4380
 Db 4761 GATGCACTAGGTATAGTACAGAGTGAATGAGTCAAGCAAGCGGCTCTGGTAGTGTG 4820
 Qy 4381 ACTGCGCAACCTCAAGCAGTGTGATATCACTGGGATTTAAACACAGTAAATGGTTA 4440
 Db 4821 ACTGCGCAACCTCAAGCAGTGTGATATCACTGGGATTTAAACACAGTAAATGGTTA 4880
 Qy 4441 AATATCATTTTCAAGAGTGGTAGAAACACTGTGCGCTTAAAGAGCAAGAAATTCAGCTG 4500
 Db 4881 AATATCATTTTCAAGAGTGGTAGAAACACTGTGCGCTTAAAGAGCAAGAAATTCAGCTG 4940

Qy 4501 AAATATATCCAGCCAGCTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACGCTCCTT 4560
 Db 4941 AAATATATCCAGCCAGCTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACGCTCCTT 5000
 Qy 4561 GAAAAGTAAAAGATTTATCTGATGAAGAAGAGAGAAACATTAGCTAAACTTGGTGTAAAGT 4620
 Db 5001 GAAAAGTAAAAGATTTATCTGATGAAGAAGAGAGAAACATTAGCTAAACTTGGTGTAAAGT 5060
 Qy 4621 GCTGTACGTTTTGTTGACCCAAATTAATACAAATTAACAGTCAATACACAAAATGAATTTACA 4680
 Db 5061 GCTGTACGTTTTGTTGACCCAAATTAATACAAATTAACAGTCAATACACAAAATGAATTTACA 5120
 Qy 4681 ACCAGACCGCTCAAGTCAAGTGAATTTCTGCAAGTAAAGCGCTGTTCTCAAGTGGTAAT 4740
 Db 5121 ACCAGACCGCTCAAGTCAAGTGAATTTCTGCAAGTAAAGCGCTGTTCTCAAGTGGTAAT 5180
 Qy 4741 GGCGCACGAGTATGTACCAATGTTGCTGACGATGGACAGCGGTAGTCAAGTAAATGACAAG 4800
 Db 5181 GGCGCACGAGTATGTACCAATGTTGCTGACGATGGACAGCGGTAGTCAAGTAAATGACAAG 5240
 Qy 4801 GTAGATTTTCATCTGCAATGAAGTCAATTTATTTTCTGATTAATTTTCTGATTAATTTT 4860
 Db 5241 GTAGATTTTCATCTGCAATGAAGTCAATTTATTTTCTGATTAATTTTCTGATTAATTTT 5300
 Qy 4861 GTTCACTACGGCTTTTACCCATCTTGTAAAAAATTTACGGAGAAATACAAATAAGTATTTT 4920
 Db 5301 GTTCACTACGGCTTTTACCCATCTTGTAAAAAATTTACGGAGAAATACAAATAAGTATTTT 5360
 Qy 4921 AACAGGTTATTTATG 4937
 Db 5361 AACAGGTTATTTATG 5377

RESULT 14
 US-08-617-697-6
 ; Sequence 6, Application US/08617697
 ; Patent No. 5977336
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/617,697
 ; FILING DATE: 01-APR-1996
 ; CLASSIFICATION: 424
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 05-OCT-1994
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-557
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9323 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-617-697-6

Query Match 97.6%; Score 4820.2; DB 2; Length 9323;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 1 TAAATATACAGATATAAATAATACAGATTTTGTGATGACAAACAACAAATTAACAA 60
 Db 441 TAAATATACAGATATAAATAATACAGATTTTGTGATGACAAACAACAAATTAACAA 500
 Qy 61 CACCTTTTTCGAGTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCCATATA 120
 Db 501 CACCTTTTTCGAGTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCCATATA 560
 Qy 121 AATGGTAAATCTTTCATCTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT 180
 Db 561 AATGGTAAATCTTTCATCTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT 620
 Qy 181 CTTTCATCTTTTCATCTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT 240
 Db 621 CTTTCATCTTTTCATCTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT 680
 Qy 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 Db 681 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
 Qy 301 GAAGCAAAATGATAAGTAATTTAAATTTTCACTTAACCTTAACCTTAACCTTAACCTTA 360
 Db 741 GAAGCAAAATGATAAGTAATTTAAATTTTCACTTAACCTTAACCTTAACCTTAACCTTA 800
 Qy 361 ATATATCGTCTCAAAATCAGCAACGCTTGAATGCTTTGGTGTCTGTCTGTAATTTGGCA 420
 Db 801 ATATATCGTCTCAAAATCAGCAACGCTTGAATGCTTTGGTGTCTGTCTGTAATTTGGCA 860
 Qy 421 CGGGTGTGTACCAATCCACAGAAAAGGCTTCGGCTATGTTACTATCTTTAGTGTATAC 480
 Db 861 CGGGTGTGTACCAATCCACAGAAAAGGCTTCGGCTATGTTACTATCTTTAGTGTATAC 920
 Qy 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGTATACATCTATTTCCA 540
 Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGTATACATCTATTTCCA 980
 Qy 541 CAATCTGTTTATGCAAGCGCTTACAGGAATGGATGTAGTACAGCGCACGCACTATG 600
 Db 981 CAATCTGTTTATGCAAGCGGCAATTTAATATGACCAACCAATGAAATGGTGCAGTTTAC 1040
 Qy 601 CAAGTAGTGGTAAATAAACCAATATCCGCAACAGTGTGACGCTATCAATTAATTTGGAA 660
 Db 1041 AAAAAACAAGTAATAAACCATTATCCGCAACAGTGTGACGCTATCAATTAATTTGGAA 1100
 Qy 661 CAATTTAAATCGACCAAAATGAATGGTGCAGTTTACAGAAAACAACAACTCCGCC 720
 Db 1101 CAATTTAAATCGACCAAAATGAATGGTGCAGTTTACAGAAAACAACAACTCCGCC 1160
 Qy 721 GTATTCAACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTATGATTTAAC 780
 Db 1161 GTATTCAACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTATGATTTAAC 1220
 Qy 781 GGACAAGTCTTTTAAATCAACCCAAATGATATCAATATGATTAAGACCAATTAATTAAC 840
 Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGATATCAATATGATTAAGACCAATTAATTAAC 1280
 Qy 841 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900
 Db 1281 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 1340

Qy 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCAGGTTTAAT 960
 Db 1341 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCAGGTTTAAT 1400
 Qy 961 ACTGTGCGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAAAGTGAAGAGGAGGTG 1020
 Db 1401 ACTGTGCGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAAAGTGAAGAGGAGGTG 1460
 Qy 1021 ATTACGCTAAATGCTGGCAGCATTTCTTTACTCGCAGGCAAAAATACCATCAGCGAT 1080
 Db 1461 ATTACGCTAAATGCTGGCAGCATTTCTTTACTCGCAGGCAAAAATACCATCAGCGAT 1520
 Qy 1081 ATATATAACCAACCACTTACTTACAGCATTTCCCGCGCTGAAATGAAGCGGTCAATCTG 1140
 Db 1521 ATATATAACCAACCACTTACTTACAGCATTTCCCGCGCTGAAATGAAGCGGTCAATCTG 1580
 Qy 1141 GCGGATATTTTTCGCAAAAGCGGTAAACATTAATGTCCGTGCTGCCACTATTCGAAACCA 1200
 Db 1581 GCGGATATTTTTCGCAAAAGCGGTAAACATTAATGTCCGTGCTGCCACTATTCGAAACCA 1640
 Qy 1201 GGTAAACTTTCCTGCTGATTCTGTAAAGCAAGATAAAAGCGCAATATTTCTTTCCGCC 1260
 Db 1641 GGTAAACTTTCCTGCTGATTCTGTAAAGCAAGATAAAAGCGCAATATTTCTTTCCGCC 1700
 Qy 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGG 1320
 Db 1701 AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGG 1760
 Qy 1321 GGCAAGCTCATGATTACAGCGGATTAAGTCAACATTAATAAAGGTGCAAGTTATCGACCT 1380
 Db 1761 GGCAAGCTCATGATTAAAGTCCGATAAAGTCAACATTAATAAAGGTGCAAGTTATCGACCT 1820
 Qy 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGAGTAAAGAAC 1440
 Db 1821 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGAGTAAAGAAC 1880
 Qy 1441 GGCATTCAATTTAGCAAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGG 1500
 Db 1881 GGCATTCAATTTAGCAAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGG 1940
 Qy 1501 AAAGAAAAAGCGGAGCGCTATTTGTGGGGGATTTGCGTTTAAATGACGCAATAT 1560
 Db 1941 AAAGAAAAAGCGGAGCGCTATTTGTGGGGGATTTGCGTTTAAATGACGCAATAT 2000
 Qy 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGTTTTGTGGAGACATCGGGCAT 1620
 Db 2001 AACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGTTTTGTGGAGACATCGGGCAT 2060
 Qy 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGAGTGGTGTAGACCTGAT 1680
 Db 2061 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGAGTGGTGTAGACCTGAT 2120
 Qy 1681 CATGTAAATTAAGCGGAGAGACCCCTTCGCAATAATACCGGTATAAATGATCAATTC 1740
 Db 2121 CATGTAAATTAAGCGGAGAGACCCCTTCGCAATAATACCGGTATAAATGATCAATTC 2180
 Qy 1741 CCACAGGACCGGTGAAGCAGCGACCTTAAAAAAATAGCGAATCAAAACACGCTA 1800
 Db 2181 CCACAGGACCGGTGAAGCAGCGACCTTAAAAAAATAGCGAATCAAAACACGCTA 2240
 Qy 1801 ACCAATACAATTAATTTCAAAATTAATCTGAAAAACGCTGGACAATGAATATACGCGATCA 1860
 Db 2241 ACCAATACAATTAATTTCAAAATTAATCTGAAAAACGCTGGACAATGAATATACGCGATCA 2300
 Qy 1861 AGAAAACTTACCGTTTAAATGCTCAATCAATCAGCAAGCAACTCCCACTTAATTTCCAT 1920
 Db 2301 AGAAAACTTACCGTTTAAATGCTCAATCAATCAGCAAGCAACTCCCACTTAATTTCCAT 2360
 Qy 1921 AGTAAAGGTGAGCGTGGCGAGCGGTTTCAAGATTCATGAGATATTTACTTTAAAGCGGA 1980
 Db 2361 AGTAAAGGTGAGCGTGGCGAGCGGTTTCAAGATTCATGAGATATTTACTTTAAAGCGGA 2420
 Qy 1981 AATTTAACCAATTTATCTGGCGGATGGGTTGATGTTTCAAAAAATATTTACCGTTGATCAG 2040

Db 2421 AATTTAAACCAATTTATCTGGCGATGGTTGATGTTTCATAAAAATATTACGCTTGATCAG 2480
QY 2041 GGTGTTTTTAATATACCGCGCTTCGTTAGCTTTTGAAGTGGAAATACAAAGCACGC 2100
Db 2481 GGTGTTTTTAATATACCGCGCTTCGTTAGCTTTTGAAGTGGAAATACAAAGCACGC 2540
QY 2101 GAGCGGCAAAATGCTAAATGCTGCGCCAGGCACTGTAAACCAATACAGGAGAGGAA 2160
Db 2541 GAGCGGCAAAATGCTAAATGCTGCGCCAGGCACTGTAAACCAATACAGGAGAGGAA 2600
QY 2161 GATTCAGGGCTAACACGCTATCTTTAAAGGAAACGGGTAAGGCTGTAATATCATTTCA 2220
Db 2601 GATTCAGGGCTAACACGCTATCTTTAAAGGAAACGGGTAAGGCTGTAATATCATTTCA 2660
QY 2221 TCAGTGAATTAATTAACCCACAATCTTAGTGCCACAAATTAACATATCTGGGAATATAA 2280
Db 2661 TCAGTGAATTAATTAACCCACAATCTTAGTGCCACAAATTAACATATCTGGGAATATAA 2720
QY 2281 ATTAACCAAACTACGAGAAAGAACCTCGTATTGGCAACACCAAGCCATGATTCGCACTGG 2340
Db 2721 ATTAACCAAACTACGAGAAAGAACCTCGTATTGGCAACACCAAGCCATGATTCGCACTGG 2780
QY 2341 AACGTCAGTGCCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2400
Db 2781 AACGTCAGTGCCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2840
QY 2401 AGCAATAGCAAAAGGCTTAACACACAGTATAGAGCTCTCGAGGGTGATTTTAAAGGC 2460
Db 2841 AGCAATAGCAAAAGGCTTAACACACAGTATAGAGCTCTCGAGGGTGATTTTAAAGGC 2900
QY 2461 GTAAATGGCAACATCTCAATCTCAAGAGAGGAGCAAGGTTAATTTCAATTAATAA 2520
Db 2901 GTAAATGGCAACATCTCAATCTCAAGAGAGGAGCAAGGTTAATTTCAATTAATAA 2560
QY 2521 CCAACGAGACATGACACAGCAACCTTTACCAATTCGGTTTATAGCCAAATATCACA 2580
Db 2961 CCAACGAGACATGACACAGCAACCTTTACCAATTCGGTTTATAGCCAAATATCACA 3020
QY 2581 GGCACGTGGGGGCTCTGTTTTTGTATATATATGCAACCAATCTTGCAGAGGGGCT 2640
Db 3021 GGCACGTGGGGGCTCTGTTTTTGTATATATATGCAACCAATCTTGCAGAGGGGCT 3080
QY 2641 GAGTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAAATTTTACCTTTAAATTCCTCAT 2700
Db 3081 GAGTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAAATTTTACCTTTAAATTCCTCAT 3140
QY 2701 GTTCGCGGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATTAATGCAACCAATTC 2760
Db 3141 GTTCGCGGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATTAATGCAACCAATTC 3200
QY 2761 AATTTTCAGCTCAGACAGAGAGATGATTTTATGACGGGTACGCAAGCAATGCAATTC 2820
Db 3201 AATTTTCAGCTCAGACAGAGAGATGATTTTATGACGGGTACGCAAGCAATGCAATTC 3260
QY 2821 AATTTCAACCTACACATATCCATCTGCGGGGTAATGTCACCCCTTGGTGACAAAATCA 2880
Db 3261 AATTTCAACCTACACATATCCATCTGCGGGGTAATGTCACCCCTTGGTGACAAAATCA 3320
QY 2881 AGCAGCAGATACGGGGATATTTACTATCGAGAAAGAGCAATATGATGCTAGAGGC 2940
Db 3321 AGCAGCAGATACGGGGATATTTACTATCGAGAAAGAGCAATATGATGCTAGAGGC 3380
QY 2941 AATAAGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGCTC 3000
Db 3381 AATAAGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGCTC 3440
QY 3001 GTTAATGGGATTTAAGTTTAACTGGGAAATGCAAGATATTAAGGCAATCTCACTATT 3060
Db 3441 GTTAATGGGATTTAAGTTTAACTGGGAAATGCAAGATATTAAGGCAATCTCACTATT 3500
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120
Db 3501 TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCTTAATATACCGGCAATTTT 3560

QY 3121 ACCAATTAAGGCACTGCCGAAATTAATATACCAAGAGAGTGGTAAACCTTTGGCAATGTT 3180
Db 3561 ACCAATTAAGGCACTGCCGAAATTAATATACCAAGAGAGTGGTAAACCTTTGGCAATGTT 3620
QY 3181 ACCAATGATGGTGAATTTAAACATTTACCACTACGCTTAAACGCAACCAAGAGCATCATC 3240
Db 3621 ACCAATGATGGTGAATTTAAACATTTACCACTACGCTTAAACGCAACCAAGAGCATCATC 3680
QY 3241 GCGGAGATATAATCAACAAAAAGGAAGCTTAAATATTACAGACAGTAATAATGATGCT 3300
Db 3681 GCGGAGATATAATCAACAAAAAGGAAGCTTAAATATTACAGACAGTAATAATGATGCT 3740
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGGCAACCTTACGATTTCTCCGAT 3360
Db 3741 GAAATCCAAATTTGGCGCAATATCTCGCAAAAAGGCAACCTTACGATTTCTCCGAT 3800
QY 3361 AAAATTAATATCAACCAACAGATAACCAATCAAAAAGGCTATTGATGGAGAGCTCTAGT 3420
Db 3801 AAAATTAATATCAACCAACAGATAACCAATCAAAAAGGCTATTGATGGAGAGCTCTAGT 3860
QY 3421 TCAGATCGGCAAGTATGCAACCTTAACCTTAATTAACCAAGAAATTTGAATTCACAGAA 3480
Db 3861 TCAGATCGGCAAGTATGCAACCTTAACCTTAATTAACCAAGAAATTTGAATTCACAGAA 3920
QY 3481 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGATTTACAGCCAAAGATGGTAGAGTTTA 3540
Db 3921 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGATTTACAGCCAAAGATGGTAGAGTTTA 3980
QY 3541 ACTATTGGCAACAGTAAATGACGGTAAACAGCGGTCGCAAGCCAAAACAGTAACCTTTAAC 3600
Db 3981 ACTATTGGCAACAGTAAATGACGGTAAACAGCGGTCGCAAGCCAAAACAGTAACCTTTAAC 4040
QY 3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCAACATGTGACACTTAAATAGCAAGTG 3660
Db 4041 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCAACATGTGACACTTAAATAGCAAGTG 4100
QY 3661 AAAACATCTAGCAGCAATGGCGACGTGAAAGCAATAGCAGCAACAGTACCGGCTTAACT 3720
Db 4101 AAAACATCTAGCAGCAATGGCGACGTGAAAGCAATAGCAGCAACAGTACCGGCTTAACT 4160
QY 3721 ATTACTCAAAAATGATAGAGTAAACAGATATTTACTTCTCTCAAAACAGTAATATC 3780
Db 4161 ATTACTCAAAAATGATAGAGTAAACAGATATTTACTTCTCTCAAAACAGTAATATC 4220
QY 3781 ACCGCTCGGAAAGGTTACCAACAGCAGCGCTCGACCATTAACGCAACAAATGGCAAA 3840
Db 4221 ACCGCTCGGAAAGGTTACCAACAGCAGCGCTCGACCATTAACGCAACAAATGGCAAA 4280
QY 3841 GCAAGTATTAACCAACAAAGGATATCAGCGGTACGATTTCCGGTTAACACCGGTAACT 3900
Db 4281 GCAAGTATTAACCAACAAAGGATATCAGCGGTACGATTTCCGGTTAACACCGGTAACT 4340
QY 3901 GTTAGCGGCACTGGTGAATTTAAACCACTTAAATCCGGCTCAAAAATTTGAAGGAAATCCGGT 3960
Db 4341 GTTAGCGGCACTGGTGAATTTAAACCACTTAAATCCGGCTCAAAAATTTGAAGGAAATCCGGT 4400
QY 3961 GAGGCTAATTAACAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTTAAATACGGTA 4020
Db 4401 GAGGCTAATTAACAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTTAAATACGGTA 4460
QY 4021 AATGTTACGCAAAAGCTGGCGATTTAAACAGTTGGGAATGGCGCAGAAATTTAAATGCGACA 4080
Db 4461 AATGTTACGCAAAAGCTGGCGATTTAAACAGTTGGGAATGGCGCAGAAATTTAAATGCGACA 4520
QY 4081 GAAGGAGCTCAACCTTAACCGCAACAGGAATACCTTGACTACTGAGCGGTTCTAGC 4140
Db 4521 GAAGGAGCTCAACCTTAACCGCAACAGGAATACCTTGACTACTGAGCGGTTCTAGC 4580
QY 4141 ATCATTCAACTAAGGTCAGGTAGACCTCTTGCTCAGAAATGGTAGCATCGCAGGAGC 4200
Db 4581 ATCATTCAACTAAGGTCAGGTAGACCTCTTGCTCAGAAATGGTAGCATCGCAGGAGC 4640

||||| 981 CAATCTGTTTGTAGCAAGCGCAATTTAACTCGACCAAAATGAATGGTGCAGTTTAC 1040
QY 601 CAAGTAGATGTAATAAACCATTATCGCAACAGTGTGACGCTATCATTAATTTGAAA 660
Db 1041 AAGAAAACAAGTAATAAACCATTATCGCAACAGTGTGACGCTATCATTAATTTGAAA 1100
QY 661 CAATTTACATCGACCAAAATGAATGGTGCAGTTTACAGAAAACAACACCTCGCC 720
Db 1101 CAATTTACATCGACCAAAATGAATGGTGCAGTTTACAGAAAACAACACCTCGCC 1160
QY 721 GATTTCAACCGTGTATACATCAACCAATTCGCCAATTTAAAGGATTTAGATTTAAC 780
Db 1161 GATTTCAACCGTGTATACATCAACCAATTCGCCAATTTAAAGGATTTAGATTTAAC 1220
QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATPAGTAAAGACGCAATTTAAC 840
Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATPAGTAAAGACGCAATTTAAC 1280
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTTACAGAAAACATCAAGCGCGCTAAT 900
Db 1281 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTTACAGAAAACATCAAGCGCGCTAAT 1340
QY 901 TTCACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAATTTGAATCAGCGTTTAAAT 960
Db 1341 TTCACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAATTTGAATCAGCGTTTAAAT 1400
QY 961 ACTGTCGGTAAAGCGGAGTGTAAATCTTATTTGGTGGCAAGTGAAMAAAGCGGGTGT 1020
Db 1401 ACTGTCGGTAAAGCGGAGTGTAAATCTTATTTGGTGGCAAGTGAAMAAAGCGGGTGT 1460
QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAATCACCATCAGCGAT 1080
Db 1461 ATTAGCGTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAATCACCATCAGCGAT 1520
QY 1081 ATAATAAACCCCAACCATTTACTTACAGCATTTGCCGGCTGAAATGAAGCGGTCAATCTG 1140
Db 1521 ATAATAAACCCCAACCATTTACTTACAGCATTTGCCGGCTGAAATGAAGCGGTCAATCTG 1580
QY 1141 GCGGATATTTTGGCAAGGCGGTAAACATTAATGTCGCTGCTGCCACATTTTCGAAACCAA 1200
Db 1581 GCGGATATTTTGGCAAGGCGGTAAACATTAATGTCGCTGCTGCCACATTTTCGAAACCAA 1640
QY 1201 GGTAACTTCTGCTGATCTGTAGCAAGATTAAGCGGCAATTTCTTTCGCGC 1260
Db 1641 GGTAACTTCTGCTGATCTGTAGCAAGATTAAGCGGCAATTTCTTTCGCGC 1700
QY 1261 AAAGAGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
Db 1701 AAAGAGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1760
QY 1321 GGCAGCTGATGATTACAGCGGATAAAGTCACATTTAAACAGGTGCGAGTTATCGACCTT 1380
Db 1761 GGCAGCTGATGATAAAGTCCGATAAAGTCACATTTAAACAGGTGCGAGTTATCGACCTT 1820
QY 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTTGGCGGTGACGCGCGGCGGAAGTAAAAAC 1440
Db 1821 TCAGGTAAAGAGGGGAGAACTTTACCTTTGGCGGTGACGCGCGGCGGAAGTAAAAAC 1880
QY 1441 GGCATTCAATTAAGCAAGAAACCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1500
Db 1881 GGCATTCAATTAAGCAAGAAACCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1940
QY 1501 AAAGAAAAGCGGAGCGCTATTCGTGGCGGATTTGGCTTAATTTGACGGCAATATT 1560
Db 1941 AAAGAAAAGCGGAGCGCTATTCGTGGCGGATTTGGCTTAATTTGACGGCAATATT 2000
QY 1561 AACGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTTTGTGGAGACATCGGGGAT 1620
Db 2001 AACGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTTTGTGGAGACATCGGGGAT 2060
QY 1621 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTCAGACCCCTGAT 1680
|||||

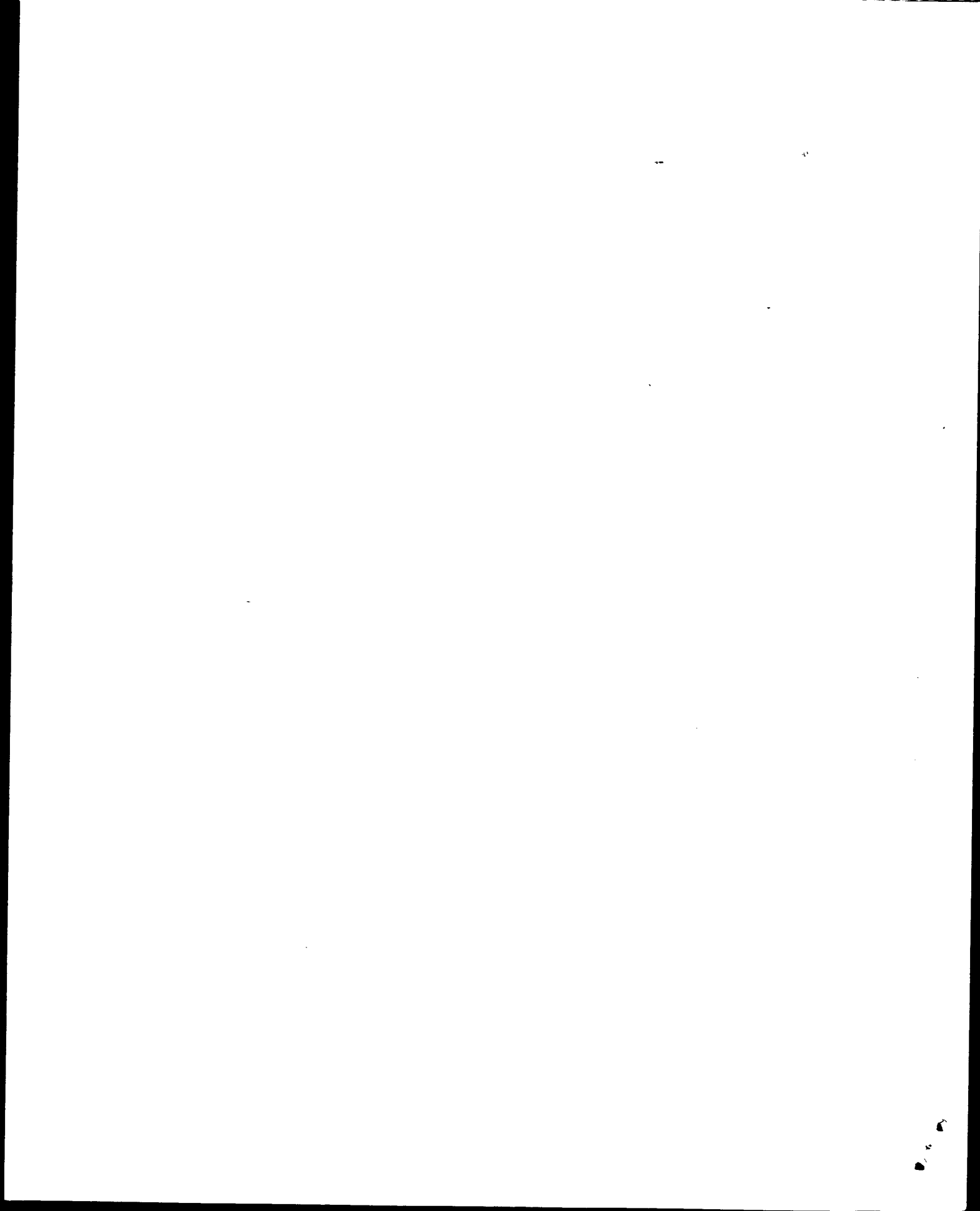
Db 2061 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTCGTAGACCCCTGAT 2120
QY 1681 GATGTAAACATTTCAAGCGGAAGACCCCTTTCCGAATTAATACCGGTATAATATGATGAATTC 1740
Db 2121 GATGTAAACATTTCAAGCGGAAGACCCCTTTCCGAATTAATACCGGTATAATATGATGAATTC 2180
QY 1741 CCACAGGACCGGTGAAGCAAGCGACCTTAAGAAAATAGCGAATCAAAACACGCTA 1800
Db 2181 CCACAGGACCGGTGAAGCAAGCGACCTTAAGAAAATAGCGAATCAAAACACGCTA 2240
QY 1801 ACCAATACAACTATTTCAAAATTTATCTGAAAAAGCGCTGGACAAATGATAAGCGCATCA 1860
Db 2241 ACCAATACAACTATTTCAAAATTTATCTGAAAAAGCGCTGGACAAATGATAAGCGCATCA 2300
QY 1861 AGAAAATTTACCCTTAATAGCTCAATCAAGCATCGGAAGCAATCCCACTTAATTTCCAT 1920
Db 2301 AGAAAATTTACCCTTAATAGCTCAATCAAGCATCGGAAGCAATCCCACTTAATTTCCAT 2360
QY 1921 AGTAAAGGTGAGGCGGTGCGGAGCGGTTCAGATTTGATGGAGATATTTCTTAAAGCGGA 1980
Db 2361 AGTAAAGGTGAGGCGGTGCGGAGCGGTTCAGATTTGATGGAGATATTTCTTAAAGCGGA 2420
QY 1981 AATTTAACCAATTTATCTGCGGATGGTGTGATTTTCAATAAAAATATTACGCTTTGATCAG 2040
Db 2421 AATTTAACCAATTTATCTGCGGATGGTGTGATTTTCAATAAAAATATTACGCTTTGATCAG 2480
QY 2041 GGTTTTTTAATATTACCGCGCTTCGTTAGCTTTTGAAGGTGGATTAACAAGCAGCG 2100
Db 2481 GGTTTTTTAATATTACCGCGCTTCGTTAGCTTTTGAAGGTGGATTAACAAGCAGCG 2540
QY 2101 GACGCGCAATGCTAAAAATTTGCGCCAGGCGACTGTAAACCATTTACAGGAGGAGAAA 2160
Db 2541 GACGCGCAATGCTAAAAATTTGCGCCAGGCGACTGTAAACCATTTACAGGAGGAGAAA 2600
QY 2161 GATTTACAGGCTAAACACGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2220
Db 2601 GATTTACAGGCTAAACACGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2660
QY 2221 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGCAATTAACATATCTGGGAATATAACA 2280
Db 2661 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGCAATTAACATATCTGGGAATATAACA 2720
QY 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGG 2340
Db 2721 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGG 2780
QY 2341 AACGTGAGTCTCTTAATCTAGACAGCGCAATTTTACCTTTTAAATACATTTCA 2400
Db 2781 AACGTGAGTCTCTTAATCTAGACAGCGCAATTTTACCTTTTAAATACATTTCA 2840
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAGCTCTCGAGGGTGAATTTTAAGCGC 2460
Db 2841 AGCAATAGCAAGGCTTAAACACACAGTATAGAGCTCTCGAGGGTGAATTTTAAGCGC 2900
QY 2461 GTAAATGGCAACATGTCATTTCAATCTCAAAAGAGGAGGAAAGTAAATTTCAATTTAAA 2520
Db 2901 GTAAATGGCAACATGTCATTTCAATCTCAAAAGAGGAGGAAAGTAAATTTCAATTTAAA 2960
QY 2521 CCAACAGAGAACATGAACACAAAGCAACCTTTACCAATTCGGTTTTTGAAGCAATATCACA 2580
Db 2961 CCAACAGAGAACATGAACACAAAGCAACCTTTACCAATTCGGTTTTTGAAGCAATATCACA 3020
QY 2581 GCCACTGGTGGGCTCTGTTTTTGGATATATATGCCAACCATTTCTGCGAGAGGGCT 2640
Db 3021 GCCACTGGTGGGCTCTGTTTTTGGATATATATGCCAACCATTTCTGCGAGAGGGCT 3080
QY 2641 GAGTTAAATAGTGAATTAATATCTTAACGGCGCTTAATTTACCTTAATTTCCAT 2700
Db 3081 GAGTTAAATAGTGAATTAATATCTTAACGGCGCTTAATTTACCTTAATTTCCAT 3140
QY 2701 GTTCGCGGATGACGCTTTTAAATCAACAAAGACTTAACCAATAAATGCAACCAATTC 2760
Db 3141 GTTCGCGGATGACGCTTTTAAATCAACAAAGACTTAACCAATAAATGCAACCAATTC 3200

us-10-092-880-3.rni

Tue Mar 25 08:39:01 2003

|||||
Db 5361 AACAGGTTATTATG 5377

Search completed: March 22, 2003, 05:46:10
Job time : 258 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 20:50:22 ; Search time 6338 Seconds
(without alignments)
12615.512 Million cell updates/sec

Title: US-10-092-880-3

Perfect score: 4937

Sequence: 1 taaatatacaagataataaa.....tttaacaggttattattatg 4937

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_hcc3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	78.2	1.6	578	17	BH000184	BH000184 2M0287B22
C 2	76.6	1.6	688	17	A2545997	A2545997 ENTEB41TR
C 3	76.2	1.5	403	17	A2654231	A2654231 1M0528K13
C 4	74.6	1.5	985	17	CNS06NAS	AL406538 T3 end of
C 5	71.2	1.4	625	17	A2605403	A2605403 1M0426P11
C 6	70.4	1.4	726	17	A2967029	A2967029 2M0237J15

7	70.4	1.4	1588	11	AK018644	Mus muscu
8	70	1.4	1101	17	CNS001T2	AL078714 Drosophil
9	69.6	1.4	490	17	BH834109	BH834109 BACFP6-K0
10	69.6	1.4	821	10	AV758661	AV758661 AV758661
11	68.8	1.4	1101	17	CNS01F77	AL107965 Drosophil
12	68.2	1.4	848	17	A2549808	A2549808 ENTEZ62TF
13	67.6	1.4	372	17	BH478403	BH478403 BGGJZ72TR
14	67.6	1.4	1063	17	CNS07A2Y	AL436064 T3 end of
15	66	1.3	342	17	AQ347002	AQ347002 RPI11-11
16	66	1.3	1036	17	CNS00599	AL057797 Drosophil
17	65.6	1.3	1101	17	CNS00EVL	AL069706 Drosophil
18	64.8	1.3	854	17	CNS0090S	AL052980 Drosophil
19	64.8	1.3	951	17	AZ676519	AZ676519 ENTGV51TF
20	64.8	1.3	1101	17	CNS0021J	AL061936 Drosophil
21	64.6	1.3	806	17	CNS04AEE	AL281759 Tetraodon
22	64.6	1.3	1101	17	CNS0039G	AL063921 Drosophil
23	64.4	1.3	437	17	CNS030FO	AL253269 Tetraodon
24	64.2	1.3	265	13	BM274994	BM274994 PfESTOaa7
25	63.6	1.3	1223	17	BL2981	BL2981 T24D11-Sp6
26	63.4	1.3	388	17	CNS0026W	AL097298 Drosophil
27	63.2	1.3	613	17	AQ222590	AQ222590 RPI1-23-2
28	63.2	1.3	1001	17	CNS0125H	AL105023 Drosophil
29	63	1.3	855	17	A2532642	AZ532642 ENTF49TR
30	63	1.3	944	17	CNS00A5G	AL054949 Drosophil
31	62.8	1.3	1055	14	BQ876453	BQ876453 AGENCOURT
32	62.8	1.3	1101	17	CNS000SU	AL075280 Drosophil
33	62.6	1.3	889	17	CNS0006W	AL065999 Drosophil
34	62.4	1.3	1885	10	BE420745	BE420745 HMM002.H0
35	62	1.3	420	17	BH094146	BH094146 RPI1-24-3
36	62	1.3	974	17	CNS001TT	AL075432 Drosophil
37	62	1.3	1101	17	CNS04TO6	AL306807 Tetraodon
38	61.8	1.3	525	10	AW906883	AW906883 EST343006
39	61.8	1.3	570	12	BG860408	BG860408 1024070D0
40	61.8	1.3	604	9	AL546530	AL546530 AL546530
41	61.8	1.3	649	13	BI920072	BI920072 EST539995
42	61.8	1.3	1059	17	CNS0022B	AL097133 Drosophil
43	61.8	1.3	1300	13	BM468018	BM468018 AGENCOURT
44	61.6	1.2	417	17	CNS02AH7	AL188548 Tetraodon
45	61.6	1.2	866	17	CNS006MS	AL065764 Drosophil

ALIGNMENTS

RESULT 1
BH000184/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH000184 578 bp DNA linear GSS 27-APR-2001
2M0287B22R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0287B22 R, DNA sequence.

BH000184
GSS.
GI:13871410

house mouse.

Mus musculus

Eukaryota; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 578)

Dunn D., Aoyagi A., Barber M., Beacorn T., Duval B., Hamil C.,

Islam H., Longacre S., Mahmoud M., Meenen E., Pedersen T., Reilly

M., Rose M., Rose R., Stokes R., Tingey A., von Niederhausen, A.

and Wright, D., Weiss, R.,

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 625)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0426 row: P column: 11
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 625.
Location/Qualifiers
1..625
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0237J15"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, P-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

FEATURES
source

1..625
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0237J15"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, P-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 311 a 63 c 161 g 90 t
ORIGIN
Query Match 1.4%; Score 71.2; DB 17; Length 625;
Best Local Similarity 75.9%; Pred. No. 2.2e-05;
Matches 88; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 126 TATAATCTTCATCTTCATCTTTAATCTTCATCTTCATCTTCATCTTC 185
Db 324 TCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 265
QY 186 ATCTTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 241
Db 264 TCT 209

RESULT 6
A2967029
LOCUS A2967029 726 bp DNA linear GSS 27-APR-2001
DEFINITION 2M0237J15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0237J15 R, DNA sequence.
ACCESSION A2967029
VERSION A2967029.1 GI:13838256

KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 726)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0237 row: J column: 15
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 726.
Location/Qualifiers
1..726
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0237J15"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, P-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

FEATURES
source

1..726
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0237J15"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, P-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 175 a 164 c 81 g 306 t
ORIGIN
Query Match 1.4%; Score 70.4; DB 17; Length 726;
Best Local Similarity 60.4%; Pred. No. 3.4e-05;
Matches 116; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 55 TTACAAACACCTTTTTCGAGCTATATGCAAAATATTTAAAAAATAGTAAATCCGCC 114
Db 396 TGAGAAATCCCTTGGTGTCTCAAGTTTACAAAGTTTAAATATTCATCTTCCTTT 455
QY 115 ATATAAATGGTATATATTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 174
Db 456 CTCCTTCT 515
QY 175 TTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 234
Db 516 TTTCCT 575

QY 235 ATCTTTCACATG 246
 Db 576 TCCTTTCCTGCTG 587

RESULT 7
 AK018644 1588 bp mRNA linear HTC 19-JAN-2002
 LOCUS Mus musculus adult male cecum CDNA, RIKEN full-length enriched
 DEFINITION library, clone:9130022E09:unclassified transcript, full insert
 sequence.
 AK018644
 VERSION AK018644.1 GI:12858457
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 PUBMED 10349636
 2
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 PUBMED 11042159
 3
 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Kashiwagi,K.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Watabiki,M.,
 Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohta,E., Watabiki,M.,
 Yoneda,Y., Ishikawa,T., Okawa,K., Tanaka,K., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multi-capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 PUBMED 11076861
 4
 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
 Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
 Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Konno,S., Yamanaka,I.,
 Saito,T., Okazaki,Y., Oshibori,T., Bono,H., Kasukawa,T., Saito,R.,
 Radota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
 Fieischmann,M., Gasterland,T., Gissi,C., King,B., Kochiwa,H.,
 Kuehl,P., Leis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
 Quackenbush,J., Schriml,L.M., Staibli,F., Suzuki,R., Tomita,M.,
 Wagner,J., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
 Baldinelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
 Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
 Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
 Marchionni,L., Mashima,J., Mazzarelli,J., Lombaerts,P., Nordone,P.,
 Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
 Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
 Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
 Wysocki,Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
 and Hayashizaki,Y.
 International annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 PUBMED 11217851
 5 (base 1 to 1588)
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
 Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
 Hara,D., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
 Hara,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
 Kawai,J., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
 Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
 Numazaki,R., Ohno,M., Okazaki,C., Sakai,K., Sano,H., Sasaki,D.,
 Saito,H., Saito,R., Sakai,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
 Schriml,L., Shibata,K., Shibata,Y., Tagami,M., Tagawa,A., Takahashi,F.,
 Sogabe,Y., Suzuki,H., Tejima,Y., Toyata,T., Yamamura,T., Yamanaka,I.,
 Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
 Hayashizaki,Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot = 10.0 and subtraction to
 Rot = 185.2. Second strand cDNA was prepared with the primer
 adapter of sequence [5'
 GAGAGAGAGATCTCGAGTTAATTAATATCCCTCCCTCCCTCC 3']. cDNA was cleaved
 with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
 bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
 end: BamHI. Host: DH10B.
 Location/Qualifiers
 1. 1588
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:9130022E09"
 /db_xref="MGD:MGI:1907754"
 /db_xref="taxon:10090"
 /clone="9130022E09"
 /sex="male"
 /tissue_type="cecum"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 1. 1588
 /note="evidence:NAS
 unclassified transcript"
 /db_xref="MGD:MGI:1921814"
 BASE COUNT 380 a 358 c 242 g 608 t
 ORIGIN
 Query Match 1.4%; Score 70.4; DB 11; Length 1588;
 Best Local Similarity 74.2%; Pred. No. 3.8e-05;
 Matches 89; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 126 TATAATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 185
 Db 602 TCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 661
 QY 186 ATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCT 245
 Db 662 CTCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTCT 721

RESULT 8
 CNS00L2/c
 LOCUS CNS00L2
 1101 bp DNA linear GSS 14-JUN-1999

DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BAC4819 of RRC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL078714
VERSION	1
KEYWORDS	AL078714.1 GI:5102004
SOURCE	GSS.
ORGANISM	Drosophila melanogaster. Drosophila melanogaster. Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101) Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieret de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RRC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
FEATURES	Location/Qualifiers
source	1..1101

	Query Match	1.4%;	Score 70;	DB 17;	Length 1101;
	Best Local Similarity	45.0%;	Pred. No. 4.4e-05;		
	Matches 108;	Conservative 35;	Mismatches 97;	Indels 0;	Gaps 0;
QY	1	TAATATACACAGATATAAAAAATAATCAAGATTTTTCTGTGATGACACAAACACAAATATTACAA	60		
Db	609	TWMMTWAMTWAAWAAWAAWATTAATAAAAWTTTWTAMTTAMTWAMHAMATWMAAWMA	550		
QY	61	CACCTTTTTTCGAGCTATATGCGCAATATTTTTAAAAAAATAGTATATAATCGCCATATAA	120		
Db	549	AAAAAAATATTAAATWAAWAAWTAATAAAAHWWTTTWMHTTTHAAAAWAAAAATWMAAAAAAAW	490		
QY	121	AATGGTATAATCTTTTCATCTTTCATCTTTAACTCTTTCATCTTTCATCTTTCATCTTTCAT	180		
Db	489	AAATTTTNTT	430		
QY	181	CTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT	240		
Db	429	TT	370		

RESULT 9	
BH834109	
LOCUS	
DEFINITION	BH834109
ACCESSION	BACPP6-K03 z Pristionchus pacificus BAC ends Pristionchus pacificus genomic DNA sequence.
VERSION	BH834109
KEYWORDS	BH834109.1 GI:21034637 GSS.
SOURCE	Pristionchus pacificus.
ORGANISM	Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus. 1 (bases 1 to 490) Srinivasan, J., Sinz, W., Lanz, C., Brand, A., Nandakumar, R., Raddatz, G., Witte, H., Keller, H., Kipping, I., Pires da Silva, A., Jesse, T., Millare, J., de Both, M., Schuster, S.C. and Sommer, R.J. A BAC-based genetic linkage map of the nematode <i>Pristionchus</i> <i>pacificus</i> Unpublished (2002) Contact: Sommer RJ Evolutionary Biology Max-Planck-Institute for Developmental Biology Spemannstr. 37-39, Tuebingen D-72076, Germany Tel: 00497071601371 Fax: 00497071601498 Email: raif.sommer@tuebingen.mpg.de Class: BAC ends.
--

FEATURES
source

```

BASE COUNT
ORIGIN
/clone_lib="Pristionchus pacificus BAC ends"
111 a 106 c 72 g 201 t

```

```
Query Match 1.48; Score 69.6; DB 17; Length 490;
Best Local Similarity 66.9%; Pred. NO. 4.9e-05;
Matches 99; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
```

91	TTAAATAAATAGTATAAAATCGCCCATATAAAATGGTATAATCTTTTCATCTTTTCATCTTTA	150
Qy		
Db		
172	TTCAAACTTTGGCTCTAATCTTTAAATCTTCATTAATCTGATCTTTACACATCTGATAGTTG	231
Qy		
151	ATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT	210
Qy		
Db		
232	ATCTTTAATCTCTGATCTTTGATCTTTTTCATCTTTTTCATCTTTGATCTTTGATCTCTGATCT	291
Qy		
211	TTTCATCTTTCATCTTTTCATCTTTTCATCT	238
Qy		
Db		
292	TTAGTCTCTGATCTCTGATCTTTTTCATCT	319
Db		

RESULT_10	AV758661/c	LOCUS	AV758661	821 bp	mRNA	linear	EST 19-OCT-2000
		DEFINITION	AV758661	BM	homo sapiens	cDNA clone BMAFA02 5',	mRNA sequence.
		ACCESSION	AV758661				
		VERSION	AV758661.1	GI:10916509			
		KEYWORDS	EST.				
		SOURCE	human.				

SOURCE ORGANISM Homo sapiens
Human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 821)

AUTHORS Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H., Gu, Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gao, G., Wang, Z., Zhang, Q., Chen, S., Han, Z. and Chen, Z.

TITLE Homo sapiens cDNA BM clones

JOURNAL Unpublished (2000)

COMMENT Contact: Zeqiang Han

FEATURES	This clone is available at CHGC in Shanghai.
source	Location/Qualifiers 1..821 /organism="Homo sapiens" /db_xref="taxon:9606"

```
Query Match 1.4% Score 68.8; DB 17; Length 1101;
Best Local Similarity 40.1%; Pred. No. 8.3e-05;
Matches 109; Conservative 51; Mismatches 112; Indels 0; Gaps 0;

/clone="BMFAEF02"
/clone_lib="BM"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/notes="Vector; pTriplex2; Site_1: sfilA; Site_2: sfilB"

BASE COUNT 311 a 139 c 111 g 257 t 3 others
ORIGIN

Query Match 1.4% Score 69.6; DB 10; Length 821;
Best Local Similarity 55.1%; Pred. No. 5.3e-05;
Matches 135; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 4 ATATCAAGATATATAAATATATCAAGATTTTGTGATGACAAACAACAAATACAAAC 63
Db 579 AAAAAAATATATAAATATATAAATATATATATATATATATATATATATATATAT 520
QY 64 CTTTTTGGAGTATATATATATATATATATATATATATATATATATATATATATAT 123
Db 519 TTTAAATATATATATATATATATATATATATATATATATATATATATATATAT 460
QY 124 GGTATATATATATATATATATATATATATATATATATATATATATATATATAT 183
Db 459 TTTTAAATATATATATATATATATATATATATATATATATATATATATATAT 400
QY 184 TCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCT 243
Db 399 ATATTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 340
QY 244 ATGAA 248
Db 339 ACAA 335

RESULT 11
LOCUS CNS017F7 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN17P12 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL107965.1 GI:5628269
VERSION AL107965.1
KEYWORDS GSS.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqreg@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
source
1..1101 Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN17P12"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/notes="end : T7"

BASE COUNT 444 a 103 c 108 g 242 t 204 others
ORIGIN

Query Match 1.4% Score 68.8; DB 17; Length 1101;
Best Local Similarity 40.1%; Pred. No. 8.3e-05;
Matches 109; Conservative 51; Mismatches 112; Indels 0; Gaps 0;

QY 113 CCATATAAATGGTATATATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCT 172
Db 670 MSTWAAAAAAATACTMTTAAACCTTTTAATWCAAMWVVVVVVVVVVVVVVVVVV 729
QY 173 TCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 232
Db 730 TTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT 789
QY 233 TCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 292
Db 790 TTTTAAWAAAAAATCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 849
QY 293 AGCTGAACCAACCAATGATTAATTAATTTTTCATCTTTCATCTTTCATCTTTCATCT 352
Db 850 SSAASASASACSAACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 909
QY 353 TGAACAGATATATCGTCTCTCAAAATTCAGCAAA 384
Db 910 AAAAAAASASASASASASASASASASASASASASASASASASASASASASAS 941

RESULT 12
LOCUS A2549808 848 bp DNA linear GSS 14-NOV-2000
DEFINITION ENTFF262TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic DNA sequence.
ACCESSION A2549808.1 GI:11174572
VERSION A2549808.1
KEYWORDS GSS.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 19
High quality sequence stop: 421.
Location/Qualifiers
1..848
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector; pIOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```



```

Db 949 AWTATTATTTTWTWTAAWAAATAAAATTAAMWATAWTAATAAAAAAATATAA 890
Qy 122 ATGGTATAATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATC 181
  || : || || || || || || || || || || || || || || || || ||
Db 889 ATWTAWAAAAATTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 830
  || || || || || || || || || || || || || || || || || || ||
Qy 182 TTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 241
  || || || || || || || || || || || || || || || || || || ||
Db 829 TTTTTCCTTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTT 770
  || || || || || || || || || || || || || || || || || || ||
Qy 242 ACATGAAT 250
  || || || || || || || || || || || || || || || || || || ||
Db 769 AAAAMAAH 761

RESULT 15
AQ347002/c
LOCUS      342 bp      DNA      linear      GSS 07-MAY-1999
DEFINITION  RPC11-116J22-TV RPC1-11 Homo sapiens genomic clone RPC1-11-116J22,
              DNA sequence.
ACCESSION  AQ347002
VERSION    AQ347002.1 GI:4171898
KEYWORDS  GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 342)
AUTHORS    Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter
            , J.C.
TITLE      Use of BAC End Sequences from Library RPC1-11 for Sequence-Ready
            Map Building
JOURNAL    Unpublished (1997)
COMMENT    Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetigr.org
            Clones are derived from the human BAC library RPC1-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES   Location/Qualifiers
            1..342
             /organism="Homo sapiens"
             /db_xref="GDB:7544397"
             /db_xref="taxon:9606"
             /clone="RPC1-11-116J22"
             /clone_lib="RPC1-11"
             /sex="Male"
             /cell_type="Lymphocytes"
             /note="Vector: pBACe3 6; Site_1: EcoRI; Site_2: EcoRI;
            RPC111 Human Male BAC Library"
BASE COUNT 158 a      6 c      31 g      147 t
ORIGIN

Query Match      1.3%; Score 66; DB 17; Length 342;
Best Local Similarity 59.7%; Pred. No. 0.00031;
Matches 111; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 65 TTTTTCGAGTCTATATGCAAAATATTTTAAAAAAATAGTAAATCCGCCATATAATG 124
  || || || || || || || || || || || || || || || || || || ||
Db 316 TATATACATATAATATATCATATATATATATATATATATATATATATATATATATC 257
  || || || || || || || || || || || || || || || || || || ||
Qy 125 GTATAATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC 184
  || || || || || || || || || || || || || || || || || || ||
Db 256 ATATCATATATCATATATATATATATATATATATATATATATATATATATATATAT 197
  || || || || || || || || || || || || || || || || || || ||

```

```

Qy 185 CATCTTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATC 244
  || || || || || || || || || || || || || || || || || || ||
Db 196 CATATATCATATATATATATATATATATATATATATATATATATATATATATAT 137
  || || || || || || || || || || || || || || || || || || ||
Qy 245 TGAAT 250
  || || || || || || || || || || || || || || || || || || ||
Db 136 TATCAT 131
  || || || || || || || || || || || || || || || || || || ||

```

Search completed: March 22, 2003, 05:41:59
Job time : 6368 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:20:59 ; Search time 46 Seconds
(without alignments)
4278.504 Million cell updates/sec

Title: US-10-092-880-4
Perfect score: 7407
Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	7389	99.8	1477	14 AAR41724	High molecular wei
2	7336	99.0	1477	18 AAW30294	Non-typeable Haemo
3	7261	98.0	1477	21 AAB01848	Haemophilus influe
4	7252	97.9	1477	15 AAR63506	Haemophilus high m
5	7193	97.1	1477	14 AAR41726	High molecular wei
6	5284	71.3	1601	18 AAW30292	Non-typeable Haemo
7	5164	69.7	1036	21 AAB01849	Haemophilus influe
8	4967	67.1	1536	14 AAR41723	High molecular wei
9	4957	66.9	1529	14 AAR41732	High molecular wei
10	4956	66.9	1536	18 AAW30293	Non-typeable Haemo

11	4943	66.7	1536	15 AAR63505	Haemophilus high m
12	4943	66.7	1536	14 AAB01846	Haemophilus influe
13	4924	66.5	1536	11 AAR41725	High molecular wei
14	4748.5	64.1	1598	18 AAR63506	Non-typeable Haemo
15	3525.5	47.6	1338	14 AAR41731	High molecular wei
16	3361.5	45.4	963	21 AAB01838	Haemophilus influe
17	3330.5	45.0	957	21 AAB01839	Haemophilus influe
18	3260	44.0	998	21 AAB01842	Haemophilus influe
19	3256.5	44.0	1079	21 AAB01836	Haemophilus influe
20	3229	43.6	992	21 AAB01843	Haemophilus influe
21	3225.5	43.5	1073	21 AAB01837	Haemophilus influe
22	3184	43.0	975	21 AAB01826	Haemophilus influe
23	3153	42.6	969	21 AAB01827	Haemophilus influe
24	2970.5	40.1	1011	21 AAB01832	Haemophilus influe
25	2939.5	39.7	1005	21 AAB01833	Haemophilus influe
26	2864	38.7	1010	21 AAB01840	Haemophilus influe
27	2833	38.2	1004	21 AAB01841	Haemophilus influe
28	2763	37.3	1095	21 AAB01847	Haemophilus influe
29	2741	37.0	1101	21 AAB01834	Haemophilus influe
30	2710	36.6	1095	21 AAB01835	Haemophilus influe
31	2678	35.7	1221	21 AAB01824	Haemophilus influe
32	2647	35.2	1227	21 AAB01825	Haemophilus influe
33	2609.5	35.2	1180	21 AAB01845	Haemophilus influe
34	2609.5	35.2	1188	21 AAB01844	Haemophilus influe
35	2592.5	35.0	1228	21 AAB01828	Haemophilus influe
36	2561.5	34.6	1222	21 AAB01830	H. influenzae stra
37	539	7.3	320	21 AAB01829	Haemophilus influe
38	517.5	7.0	1978	20 AAY27230	Amino acid sequenc
39	508	6.9	314	21 AAB01831	Haemophilus influe
40	508	6.9	1981	19 AAR42634	Protein sequence t
41	498.5	6.7	2411	21 AAB33860	Haemophilus influe
42	491	6.6	2353	17 AAR93933	Haemophilus adhesi
43	489	6.6	2123	22 AAR00701	Moraxella catarrha
44	489	6.6	3241	22 AAG66005	F. necrophorum leu
45	476	6.4	1532	20 AAY27231	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAR41724
ID AAR41724 standard; Protein; 1477 AA.
XX
AC AAR41724;
XX
DT 26-APR-1994 (first entry)
XX
DE High molecular weight protein 2 (HMW2).
XX
KW HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX
OS Haemophilus influenzae.
XX
PN WO9319090-A.
XX
PD 30-SEP-1993.
XX
PF 16-MAR-1993; 93WO-0502166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
(BARE/) BARENKAMP S J.
(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Barenkamp SJ;
XX
DR WPI: 1993-320683/40.
XX
N-PSDB; AAQ49507.
XX
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties

xx
ps
xx
cc
cc
cc
cc
cc
cc
cc
cc
sq

Claim 4; Figure 4; 100pp; English.

The isolation and purification of the high molecular weight protein enables the identification of the major protective epitopes of the protein by conventional epitope mapping. These epitopes can then be synthesised using standard techniques and incorporated into fully synthetic or recombinant vaccines.

Query Match 99.8%; Score 7389; DB 14; Length 1477;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKSEKPKARKVRLHAKLPKLSAMLLSLGVT 60
DB 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKSEKPKARKVRLHAKLPKLSAMLLSLGVT 60

QY 61 SIPOSVLASGLQGMVHGVTATMQVDCNKTIIIRNSVDALINWKQFNIDQNMVQFLOENN 120
DB 61 SIPOSVLASGLQGMVHGVTATMQVDCNKTIIIRNSVDALINWKQFNIDQNMVQFLOENN 120

QY 121 NSAFNVRVTSNQISQLKGLDSNGQVFLNPNGITIGKDALINTNGFTASTLDISNENIK 180
DB 121 NSAFNVRVTSNQISQLKGLDSNGQVFLNPNGITIGKDALINTNGFTASTLDISNENIK 180

QY 181 ARNFTFQTKKALAEIVNHLITVKGDSVNLIGGKVKNEGIVSVNGSSISLLAGQKIT 240
DB 181 ARNFTFQTKKALAEIVNHLITVKGDSVNLIGGKVKNEGIVSVNGSSISLLAGQKIT 240

QY 241 ISDIINPTITYSIAAPENAVNLGDIIPAKGGINVRATIRNOGKLSADSVSKDKSNIV 300
DB 241 ISDIINPTITYSIAAPENAVNLGDIIPAKGGINVRATIRNOGKLSADSVSKDKSNIV 300

QY 301 LSAKEGEAIGVISAQNOQAKGKLMITGDKVTLKTGAVIDLSGKEGGETVYLGDERGE 360
DB 301 LSAKEGEAIGVISAQNOQAKGKLMITGDKVTLKTGAVIDLSGKEGGETVYLGDERGE 360

QY 361 GKNGIOLAKKTSLEKGSTINVSKEGKGFPAIVMGDIALIDGNINAQSGCDIAKGGFVET 420
DB 361 GKNGIOLAKKTSLEKGSTINVSKEGKGFPAIVMGDIALIDGNINAQSGCDIAKGGFVET 420

QY 421 SGHDLFIKNAIVDAKEWLLDFDNVINAEDPLRNNTGINDPEPTGTGEASDPKKNSLKL 480
DB 421 SGHDLFIKNAIVDAKEWLLDFDNVINAEDPLRNNTGINDPEPTGTGEASDPKKNSLKL 480

QY 481 TLTNTTISNLYKNAWTMNTITASRKLTVNNSINIGSNLSHLILHSKGQGGVQIDGDIITS 540
DB 481 TLTNTTISNLYKNAWTMNTITASRKLTVNNSINIGSNLSHLILHSKGQGGVQIDGDIITS 540

QY 541 KGNLTYSGGWVDVHKNIPLDQGFNLITAAVAFEGGNKARDAANAIVAOQTVTITG 600
DB 541 KGNLTYSGGWVDVHKNIPLDQGFNLITAAVAFEGGNKARDAANAIVAOQTVTITG 600

QY 601 EGKDFRANVSLNGTQKGLNLTSSVNNLTNLSCTINISGNITINQTRKNTSYWQTSHD 660
DB 601 EGKDFRANVSLNGTQKGLNLTSSVNNLTNLSCTINISGNITINQTRKNTSYWQTSHD 660

QY 661 SHWNVSALNLETGANFTFIKYISSNSKGLTTQYRSSAGVNFNGVNMNFKLKEGAKVNF 720
DB 661 SHWNVSALNLETGANFTFIKYISSNSKGLTTQYRSSAGVNFNGVNMNFKLKEGAKVNF 720

QY 721 KUKPENNMNTSKPLPIREFLANITATGGSVFFDIYANHSGRGAELKWEINISNGANFTL 780
DB 721 KUKPENNMNTSKPLPIREFLANITATGGSVFFDIYANHSGRGAELKWEINISNGANFTL 780

QY 781 NSHVRGDDAFKINKDLTINATNSNFSRQTKDDFDYGYARNAINSTYINISILGNVTLGG 840
DB 781 NSHVRGDDAFKINKDLTINATNSNFSRQTKDDFDYGYARNAINSTYINISILGNVTLGG 840

QY 841 QNSSSITGNITTEKAAVNTLEANNAPQONIDRVIKLGSLLVNGSLSTGTGNADIKGN 900
DB 841 QNSSSITGNITTEKAAVNTLEANNAPQONIDRVIKLGSLLVNGSLSTGTGNADIKGN 900

DB 841 QNSSSITGNITTEKAAVNTLEANNAPQONIDRVIKLGSLLVNGSLSTGTGNADIKGN 900

QY 901 LTISESATFKGTRDTLNTGNFTNNGTAETINITQGVVKLGNTDGDNLNITTHAKRNQR 960
DB 901 LTISESATFKGTRDTLNTGNFTNNGTAETINITQGVVKLGNTDGDNLNITTHAKRNQR 960

QY 961 SLIIGCDIINKKGSNLNTSDNDAEIQIGGNIQSKKEGNTLTISSDKINITQIITKKGIDGE 1020
DB 961 SLIIGCDIINKKGSNLNTSDNDAEIQIGGNIQSKKEGNTLTISSDKINITQIITKKGIDGE 1020

QY 1021 DSSSDATSNANLTITKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV 1080
DB 1021 DSSSDATSNANLTITKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV 1080

QY 1081 TFNNVKDSKISADGHNVTLNSKVKTSSSNGGRESNDNDTGLTITAKNVEVKNKDTISLKT 1140
DB 1081 TFNNVKDSKISADGHNVTLNSKVKTSSSNGGRESNDNDTGLTITAKNVEVKNKDTISLKT 1140

QY 1141 VNITASEKVTTTAGTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200
DB 1141 VNITASEKVTTTAGTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200

QY 1201 KSGEANVTSATGTIGGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
DB 1201 KSGEANVTSATGTIGGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260

QY 1261 GSSITSTKQVDDLLAQNQSGTAGSINAAVNTLNTGTLTTVAGSDIKATSGTLVINAKDAK 1320
DB 1261 GSSITSTKQVDDLLAQNQSGTAGSINAAVNTLNTGTLTTVAGSDIKATSGTLVINAKDAK 1320

QY 1321 LNGLDASGDSSTEVNAVNASGSGSVTAATSSSVNITGDLNTVNGLNITISKDGNTVRLRGKE 1380
DB 1321 LNGLDASGDSSTEVNAVNASGSGSVTAATSSSVNITGDLNTVNGLNITISKDGNTVRLRGKE 1380

QY 1381 IEVKYIQPCVASVEVTEAKRVLEKVKOLSDERETLAKLGVSARVFEVPEPNTITVNTQN 1440
DB 1381 IEVKYIQPCVASVEVTEAKRVLEKVKOLSDERETLAKLGVSARVFEVPEPNTITVNTQN 1440

QY 1441 EFTTRPSSQVIIESEKACFSSGNGARVCTNVAADDGP 1477
DB 1441 EFTTRPSSQVIIESEKACFSSGNGARVCTNVAADDGP 1477

RESULT 2
AAW30294
ID AAW30294 standard; Protein; 1477 AA.
XX
AC AAW30294;
XX
DT 14-APR-1998 (first entry)
XX
DE Non-typeable Haemophilus high mol.wt. surface protein HMW2.
XX
KW Non-typeable Haemophilus; high molecular weight surface protein;
KW HMW2; hmw2A gene; immunogen; vaccine; otitis media.
XX
OS Haemophilus influenzae strain 12.
XX
PH Key Location/Qualifiers
FT Misc-difference 34 /note= "encoded by TTC"
FT Misc-difference 35 /note= "encoded by CGC"
FT Misc-difference 36 /note= "encoded by TAT"
FT Misc-difference 37 /note= "encoded by GTT"
FT Misc-difference 38 /note= "encoded by ACT"
FT Misc-difference 39 /note= "encoded by ATC"
FT Misc-difference 40 /note= "encoded by TTT"

FT	Misc-difference 41	/note=	"encoded by AGG"	Db	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTKESRYVTIFRCNHLALPLSAMLISLGV	60
FT	Misc-difference 42	/note=	"encoded by TGT"	Qy	61	SIPOSVLASGLOGMDVHGHTATMOVDGKNTIIRNSVDALINNKOFNIDONEMVOFLOENN	120
FT	Misc-difference 43	/note=	"encoded by AAC"	Db	61	SIPOSVLASGLOGMDVHGHTATMOVDGKNTIIRNSVDALINNKOFNIDONEMVOFLOENN	120
FT	Misc-difference 424	/note=	"encoded by TAT"	Qy	121	NSAVFNRTVTSNQISQLKGLDSNGQVFLINPGITIGKDAIINTNGFTASTLIDISNENIK	180
FT	Misc-difference 426	/note=	"encoded by TCC"	Db	121	NSAVFNRTVTSNQISQLKGLDSNGQVFLINPGITIGKDAIINTNGFTASTLIDISNENIK	180
FT	Misc-difference 428	/note=	"encoded by GAC"	Qy	181	ARNFTFEOTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSGISILLAGCKIT	240
FT	Misc-difference 429	/note=	"encoded by AGC"	Db	181	ARNFTFEOTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSGISILLAGCKIT	240
FT	Misc-difference 434	/note=	"encoded by AAA"	Qy	241	ISDIINPTITYSTAAPENEAVNLGDIFAKGGNINVRAATIRNOGKLSADSVSKDKSGNIV	300
FT	Misc-difference 435	/note=	"encoded by ACA"	Db	241	ISDIINPTITYSTAAPENEAVNLGDIFAKGGNINVRAATIRNOGKLSADSVSKDKSGNIV	300
FT	Misc-difference 442	/note=	"encoded by CCT"	Qy	301	LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGETVILGGDERGE	360
FT	Misc-difference 444	/note=	"encoded by GAT"	Db	301	LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGETVILGGDERGE	360
FT	Misc-difference 446	/note=	"encoded by ACA"	Qy	361	GNKGIOLAKKTSLEKGSTINVSKEKEGGFAIVMGDIALIDGNINAGSGDIAKTGGFVET	420
FT	Misc-difference 448	/note=	"encoded by GAA"	Db	361	GNKGIOLAKKTSLEKGSTINVSKEKEGGFAIVMGDIALIDGNINAGSGDIAKTGGFVET	420
FT	Misc-difference 454	/note=	"encoded by CGC"	Qy	421	SGHDLFIKDNAIVDAKEWLLDFDNVSIINAEPLFNNTGINDEPTGTGEASDPKKNSELK	480
FT				Db	421	SGHDLFIKDNAIVDAKEWLLDFDNVSIINAEPLFNNTGINDEPTGTGEASDPKKNSELK	480
PN	WO9736914-A1.			Qy	481	TTLTNTTISNYLKNWMTNITASRKLTIVNSSINIGNSHLILHSGKQGGGVQIDGDITS	540
XX	09-OCT-1997.			Db	481	TTLTNTTISNYLKNWMTNITASRKLTIVNSSINIGNSHLILHSGKQGGGVQIDGDITS	540
XX	01-APR-1997; 97WO-US04707.			Qy	541	KGKGLTIYSGGWVDVHKNTITLDQGFNLITAAASVAFEGGNKARDAANAKIAVAGCTVTTIG	600
XX	01-APR-1996; 96US-0617697.			Db	541	KGKGLTIYSGGWVDVHKNTITLDQGFNLITAAASVAFEGGNKARDAANAKIAVAGCTVTTIG	600
PA	(BARE/) BARENKAMP S J.			Qy	601	EGKDFRANNVSLNGTGKGLNIISSVNNLTHNLSGTTINISGNTITINOTTRKNTSYWQTSHD	660
PI	Barenkamp SJ;			Db	601	EGKDFRANNVSLNGTGKGLNIISSVNNLTHNLSGTTINISGNTITINOTTRKNTSYWQTSHD	660
DR	WPI; 1997-503038/46.			Qy	661	SHWVNSALNLETGANFTFIKYIISNSKGLTQYRSSAGVNFNGVNGMSEFNKEGAKVNF	720
DR	N-PSDB; AAT90995.			Db	661	SHWVNSALNLETGANFTFIKYIISNSKGLTQYRSSAGVNFNGVNGMSEFNKEGAKVNF	720
PT	High molecular weight proteins of non-typeable Haemophilus			Qy	721	KLKPNENMNTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMSSEINISGANFTL	780
PT	influenzae - useful for vaccine production			Db	721	KLKPNENMNTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMSSEINISGANFTL	780
PS	Claim 7; Page 73-78; 183pp; English.			Qy	781	NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNTISILGNNVTIIG	840
XX	This protein comprises the high molecular weight surface protein			Db	781	NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNTISILGNNVTIIG	840
CC	HMW2 (123 kDa) of non-typeable Haemophilus influenzae strain 12 that			Qy	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDRVTKLGSLLVNGSLSTGENADIKGN	900
CC	has the immunological ability to protect against disease caused by a			Db	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDRVTKLGSLLVNGSLSTGENADIKGN	900
CC	non-typeable Haemophilus strain and is characterised by at least			Qy	901	LTISATFPGKTRDTLNTITGNTNGTAEINTQGVVVKLGNTVNDGDLNITTHAKRNOR	960
CC	one surface-exposed B-cell epitope that is recognised by monoclonal			Db	901	LTISATFPGKTRDTLNTITGNTNGTAEINTQGVVVKLGNTVNDGDLNITTHAKRNOR	960
CC	antibody AD6. The HMW2 amino acid sequence was deduced from the			Qy	961	SIIGDDIINKKGLSINITDSNDAEITQGGNISQEGNLTISDDKINTTKQITIKKIDGGE	1020
CC	hmw2 gene sequence (see AAT90995 and AAT90997). The expressed protein			Db	961	SIIGDDIINKKGLSINITDSNDAEITQGGNISQEGNLTISDDKINTTKQITIKKIDGGE	1020
CC	is truncated, starting at residue 442 of the full-length gene			Qy	1021	DSSSDATSNANLTITKTELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV	1080
CC	product. HMW1 (see AAW30293), HMW3 (see AAW30291) and HMW4 (see			Db	1021	DSSSDATSNANLTITKTELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV	1080
CC	AAW30292) have also been identified. A conjugate comprising HMW2			Qy	1081	TFNNVKOSKISADGHNVTLSKVKTSNNGRESNDSNDTGLTITAKNVEVNDKIDISLKT	1140
CC	linked to an antigen, haptan or polysaccharide, and a synthetic			Db	1081	TFNNVKOSKISADGHNVTLSKVKTSNNGRESNDSNDTGLTITAKNVEVNDKIDISLKT	1140
CC	peptide of 6-150 amino acids corresponding to at least protective			Qy			
CC	epitope of HMW2 are also claimed. HMW proteins, conjugates and			Db			
CC	peptides can be used in vaccines, as immunogens for preparation of			Qy			
CC	antibodies and as antigens for detection of these antibodies.			Db			
XX	Sequence 1477 AA;			Qy			
XX	Query Match 99.0%; Score 7336; DB 18; Length 1477;			Db			
XX	Best Local Similarity 99.2%; Pred. No. 0;			Qy			
XX	Matches 1465; Conservative 1; Mismatches 11; Indels 0; Gaps 0;			Db			
QY	1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKESRYVTIFRCNHLALPLSAMLISLGV			Qy			
				Db			

QY 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEA 1200
 Db 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEA 1200
 QY 1201 KSGEANTVATGTTIGTISGNTVNTANAGDLTVNGAEINATGATLTATGNTLTTEA 1260
 Db 1201 KSGEANTVATGTTIGTISGNTVNTANAGDLTVNGAEINATGATLTATGNTLTTEA 1260
 QY 1261 GSSITSTKGVLDLQAQSGSIAGSINAANVTNTGTLTFTVAGSDIKATGTLVINAADAK 1320
 Db 1261 GSSITSTKGVLDLQAQSGSIAGSINAANVTNTGTLTFTVAGSDIKATGTLVINAADAK 1320
 QY 1321 LNGDASGDSFEVNAVNASGSGSVTAATSSSVNITGDLTVNGNLISKDRNTVRLRKE 1380
 Db 1321 LNGDASGDSFEVNAVNASGSGSVTAATSSSVNITGDLTVNGNLISKDRNTVRLRKE 1380
 QY 1381 IEVKYIQPGVASVEEIVIAKRVLEKVKDLSDERETLAKLGVSAVRFVPPNNTITVNTQN 1440
 Db 1381 IEVKYIQPGVASVEEIVIAKRVLEKVKDLSDERETLAKLGVSAVRFVPPNNTITVNTQN 1440
 QY 1441 EFTTRPSSQVILISEGKACFSGNGARVCTNVADGQP 1477
 Db 1441 EFTTRPSSQVILISEGKACFSGNGARVCTNVADGQP 1477

RESULT 3

AAB01848
 ID AAB01848 standard; Protein: 1477 AA.

AC AAB01848;

DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain 12 HMW2A protein, SEQ ID NO:71.

XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis.

OS Haemophilus influenzae strain 12.

XX WO200020609-A2.

PN 13-APR-2000.

XX 07-OCT-1999; 99WO-CA00938.

PR 07-OCT-1998; 98US-0167568.

PR 08-DEC-1998; 98US-0206942.

XX (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Yang Y, Klein MH;

XX WPI: 2000-303789/26.

DR N-PSDB; AA52197.

XX Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -

XX Example 16; Fig 29A-N; 307pp; English.

XX The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,

CC and the hmwB and hmwC genes encode the structural HMW proteins
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMW. The invention also discloses hmwA genes (AA52175-AA52198)
 CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
 CC strains J95c, K1, K21, L20C2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMW protein from a non-typeable strain of
 CC H. influenzae.

XX Sequence 1477 AA;

Query Match 98.0%; Score 7261; DB 21; Length 1477;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1454; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSEKPARMKVPHLALPLSAMLISLGV 60
 Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSEKPARMKVPHLALPLSAMLISLGV 60

QY 61 9IPQSVLASGLQGMVDVVGHTATMQVDGKNTIIRNSVDALINMKQFNIDQNMVQFLOENN 120
 Db 61 9IPQSVLASGLQGMVDVVGHTATMQVDGKNTIIRNSVDALINMKQFNIDQNMVQFLOENN 120

QY 121 NSAVFNRTVSNQISQLGILDSNGOVFLINPNTIGTKDAINTNGFTASTLDLSNENIK 180
 Db 121 NSAVFNRTVSNQISQLGILDSNGOVFLINPNTIGTKDAINTNGFTASTLDLSNENIK 180

QY 181 ARNFTFEQTKDALKAEIVNHGLITVKGDSVNLIGGKVKNEGIVSVNGSGISLLAGOKIT 240
 Db 181 ARNFTFEQTKDALKAEIVNHGLITVKGDSVNLIGGKVKNEGIVSVNGSGISLLAGOKIT 240

QY 241 ISDIINPTITYSIAAPENEAVALNGDIFAKGGINVRAATIRNOGKLSDSVSKDKSGNIV 300
 Db 241 ISDIINPTITYSIAAPENEAVALNGDIFAKGGINVRAATIRNOGKLSDSVSKDKSGNIV 300

QY 301 LSAKGEAEIGGVISAQNOQAQGGKLMITGDKVTLKTGAVIDLSGKEGETYLGGERGE 360
 Db 301 LSAKGEAEIGGVISAQNOQAQGGKLMITGDKVTLKTGAVIDLSGKEGETYLGGERGE 360

QY 361 GKNGIQLAKTSLKSGSTINVSQKKGFAIWMGDIALDGNINAQSGSDIAKTGGFVET 420
 Db 361 GKNGIQLAKTSLKSGSTINVSQKKGFAIWMGDIALDGNINAQSGSDIAKTGGFVET 420

QY 421 SGHDLFIKDNAIVDAKEWLLDFDNVSVINAEDEPLFNNTGINDPEPTGTEASDPKKNSL 480
 Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSVINAEDEPLFNNTGINDPEPTGTEASDPKKNSL 480

QY 481 TLTNTTISNLYKNAMTMTITASRKLTVNSSINIGSNHLLIHSKGGGQVQIDGDI 540
 Db 481 TLTNTTISNLYKNAMTMTITASRKLTVNSSINIGSNHLLIHSKGGGQVQIDGDI 540

QY 541 KGNLTITISGGWVDVHKNTITLDQGFNLITAAVAFEGGNKARDAAAKIVAGQTVITG 600
 Db 541 KGNLTITISGGWVDVHKNTITLDQGFNLITAAVAFEGGNKARDAAAKIVAGQTVITG 600

QY 601 ECKDFRANNVSLNGTCKGLNISSVNNLTNLSGNTINISGNTINQTRKNTSYWQTS 660
 Db 601 ECKDFRANNVSLNGTCKGLNISSVNNLTNLSGNTINISGNTINQTRKNTSYWQTS 660

QY 661 SHWNYSALNLETGANTFTFKYISSNSKGLTQYRSSAGVNFVNGVNGNSFNLEKAGVNF 720
 Db 661 SHWNYSALNLETGANTFTFKYISSNSKGLTQYRSSAGVNFVNGVNGNSFNLEKAGVNF 720

QY 721 KLPENNMSTKPLPIRFLANITATGGSVFEDYIYANHSGRAELKMSSEINISNGANFTL 780
 DB 721 KLPENNMSTKPLPIRFLANITATGGSVFEDYIYANHSGRAELKMSSEINISNGANFTL 780
 QY 781 NSHVGGDAFKINKDLTINATNSFSLRQTKDDFYDYARNAINSTYINISILGNNVTGLG 840
 DB 781 NSHVGGDAFKINKDLTINATNSFSLRQTKDDFYDYARNAINSTYINISILGNNVTGLG 840
 QY 841 QNSSSSITGNTTIEKAANVTLEANNAPNOQNIQIRVIRKLGSLVNGSLTGENADIKGN 900
 DB 841 QNSSSSITGNTTIEKAANVTLEANNAPNOQNIQIRVIRKLGSLVNGSLTGENADIKGN 900
 QY 901 LTISATFFKGTTRDTLNTGNTNGTAEINITQGVVVKLGNTVNDGDLNITTHAKRNOR 960
 DB 901 LTISATFFKGTTRDTLNTGNTNGTAEINITQGVVVKLGNTVNDGDLNITTHAKRNOR 960
 QY 961 SIIGDIIKKGSLNITDSNNDAEIQIGGNISQKEGNLTITSSDKINITKQITIKKGDGE 1020
 DB 961 SIIGDIIKKGSLNITDSNNDAEIQIGGNISQKEGNLTITSSDKINITKQITIKKGDGE 1020
 QY 1021 DSSSDATSNANLTITKTELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGSCAEAKTV 1080
 DB 1021 DSSSDATSNANLTITKTELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGSCAEAKTV 1080
 QY 1081 TFNNVKDSKISADGHNVTLSKVTSSNGGREGSNDNDTGLTITAKNVEVNDKIDTSLKT 1140
 DB 1081 TFNNVKDSKISADGHNVTLSKVTSSNGGREGSNDNDTGLTITAKNVEVNDKIDTSLKT 1140
 QY 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTTKSGSKIEA 1200
 DB 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTTKSGSKIEA 1200
 QY 1201 KSGEANVTSAITGTTGGTISGNTVNTVNTAGDLTVGNGAEINATEGAATLTATGNTLITEA 1260
 DB 1201 KSGEANVTSAITGTTGGTISGNTVNTVNTAGDLTVGNGAEINATEGAATLTATGNTLITEA 1260
 QY 1261 GSSITSTKGVQVLLAQNGSIAGSINAANVTLLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320
 DB 1261 GSSITSTKGVQVLLAQNGSIAGSINAANVTLLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320
 QY 1321 LMGDASGDSPTENVANASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTVRLRGKE 1380
 DB 1321 LMGDASGDSPTENVANASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTVRLRGKE 1380
 QY 1381 TEVKYIQPCVASVEVIEAKRVLEKVDLSDEERETLAKLGVSASVRFVEPNNTITVNTQN 1440
 DB 1381 TEVKYIQPCVASVEVIEAKRVLEKVDLSDEERETLAKLGVSASVRFVEPNNTITVNTQN 1440
 QY 1441 EFTTRPSSQVVISSEKACFSSGNGARVCTNVADGQOP 1477
 DB 1441 EFTTRPSSQVVISSEKACFSSGNGARVCTNVADGQOP 1477
 RESULT 4
 ID AAR63506 standard; Protein; 1477 AA.
 XX AAR63506;
 AC AAR63506;
 DT 25-JUN-1995 (first entry)
 XX Haemophilus high molecular weight protein HMW2.
 XX High molecular weight protein; HMW2; protective vaccine; otitis;
 KW sinusitis; bronchitis; Hib; ss.
 XX Haemophilus.
 OS WO9421290-A.
 PN 29-SEP-1994.
 XX
 PD
 XX

PF 15-MAR-1994; 94WO-US02550.
 XX
 PR 16-MAR-1993; 93US-0038682.
 XX
 PA (BARE/) BARENKAMP S J.
 PA (SGEM/) ST GEME J W.
 XX
 PI Barenkamp SJ, St GEME JW;
 XX
 XX WPI; 1994-316665/39.
 DR Q-PSDB; Q72294.
 XX
 XX New immunogenic high mol. wt. proteins of non typeable
 PT Haemophilus - useful in protective vaccines
 XX
 XX Claim 3; Page 36; 127pp; English.
 XX
 CC The HMW2 protein encoded by this sequence is useful in a vaccine to
 CC protect against disease caused by non-typeable Haemophilus which are
 CC not controlled by H. influenzae type b (Hib) vaccines. The encoded
 CC protein can also be used as a carrier for protective Hib
 CC polysaccharide (in a conjugate vaccine against meningitis) or for
 CC other antigens, haptens, etc.
 XX
 XX Sequence 1477 AA;
 SQ
 Query Match 97.9%; Score 7252; DB 15; Length 1477;
 Best Local Similarity 98.3%; Pred. No. 0;
 Matches 1452; Conservative 4; Mismatches 21; Indels 0; Gaps 0;
 QY 1 MNKIYRLKFSKRNLAVAVSELARGCDHSTEGSEKPAKMKVHRLKALKPLSAMLGLVGT 60
 DB 1 MNKIYRLKFSKRNLAVAVSELARGCDHSTEGSEKPAKMKVHRLKALKPLSAMLGLVGT 60
 QY 61 SIPOSVLASLOGMDVHVGTATMOVDGNKTIIRNSVDIAIINMKQFNIDQNMVQFLOENN 120
 DB 61 SIPOSVLASLOGMDVHVGTATMOVDGNKTIIRNSVDIAIINMKQFNIDQNMVQFLOENN 120
 QY 121 NSAVFNRTVSNQISQLKGLDSDNGQVFLINPNTGITKDAIINTNGFTASTLDSINENIK 180
 DB 121 NSAVFNRTVSNQISQLKGLDSDNGQVFLINPNTGITKDAIINTNGFTASTLDSINENIK 180
 QY 181 ARNFTFQTKDALKAEIVNHGLITVKGDSVNLIGKVKNEGVIYVNGGSIISLAGOKIT 240
 DB 181 ARNFTFQTKDALKAEIVNHGLITVKGDSVNLIGKVKNEGVIYVNGGSIISLAGOKIT 240
 QY 241 ISDIINPTTITYSTAAPENEAENVLDIFAKGGINVRAATIRNOCKLSADSVSKDKSGNIV 300
 DB 241 ISDIINPTTITYSTAAPENEAENVLDIFAKGGINVRAATIRNOCKLSADSVSKDKSGNIV 300
 QY 301 LSAKEGEAEITGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGBERGE 360
 DB 301 LSAKEGEAEITGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGBERGE 360
 QY 361 GKNGIQLAKKTSLEKSTINVSQKEKGGFAIYVWGDIALIDGNINAQSGDIAKTGGFVET 420
 DB 361 GKNGIQLAKKTSLEKSTINVSQKEKGGFAIYVWGDIALIDGNINAQSGDIAKTGGFVET 420
 QY 421 SGHDLFTKDAIYDAKEWLLDFDNVSINAEPLNFNTGTINDEPFTGTGEASDPKKNSELK 480
 DB 421 SGHDLFTKDAIYDAKEWLLDFDNVSINAEPLNFNTGTINDEPFTGTGEASDPKKNSELK 480
 QY 481 TLTNTTNTISYLNKNAWMTNITASRKLTVNSSINIGNSHLLILHSKGORGGVQIDGDIITS 540
 DB 481 TLTNTTNTISYLNKNAWMTNITASRKLTVNSSINIGNSHLLILHSKGORGGVQIDGDIITS 540
 QY 541 KGNLTIIYSGWVDVHKNTLTDGFLNITAAVSVAFFEGGNKARDAANAKIYVAGTWTITG 600
 DB 541 KGNLTIIYSGWVDVHKNTLTDGFLNITAAVSVAFFEGGNKARDAANAKIYVAGTWTITG 600
 QY 601 EGKDFRANVSLNGTCKGLNIISSVNNLTNLSGTTINISGNITINOTTRKNTSYWOTSHD 660
 DB 601 EGKDFRANVSLNGTCKGLNIISSVNNLTNLSGTTINISGNITINOTTRKNTSYWOTSHD 660


```
Db 541 KGNLTYSGGWVDVHKNTLDGFLNITTAASVAFEGGNKARDAANAQIVTITG 600
Qy 601 EGKDFRANVSLNGTGKGLNLTSSVNNLTHNLSTGTLNINSGNTITNGTTRKNTSYQWTS 660
Db 601 EGKDFRANVSLNGTGKGLNLTSSVNNLTHNLSTGTLNINSGNTITNGTTRKNTSYQWTS 660
Qy 661 SHNVSALENGTGANFTFKYISNSSKGLTQYRSSAGVNFNGVNMFSNLKEGAKVNF 720
Db 661 SHNVSALENGTGANFTFKYISNSSKGLTQYRSSAGVNFNGVNMFSNLKEGAKVNF 720
Qy 721 KLPNENMTSKPLPIRFLANITATGGSVFFDIYAHNSGRGAELKMSINISNGANFTL 780
Db 721 KLPNENMTSKPLPIRFLANITATGGSVFFDIYAHNSGRGAELKMSINISNGANFTL 780
Qy 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNIISILGNNVT 840
Db 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNIISILGNNVT 840
Qy 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIIRDVVKLGSLLVNGSLTGENADIKGN 900
Db 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIIRDVVKLGSLLVNGSLTGENADIKGN 900
Qy 901 LTISEATFKGKPTDLNLTGNTNGTAEINITQGVVKLGNTVNDGDLNITTHAKRNOR 960
Db 901 LTISEATFKGKPTDLNLTGNTNGTAEINITQGVVKLGNTVNDGDLNITTHAKRNOR 960
Qy 961 SIIGDDIINKGSLNITDSNDAEIQIGNLSQEGNLTISDQKINTKQITIKKGIDGE 1020
Db 961 SIIGDDIINKGSLNITDSNDAEIQIGNLSQEGNLTISDQKINTKQITIKKGIDGE 1020
Qy 1021 DSSSDATSNANLTITKELKLTEDLSISGNKAEITAKOGRDLTIGNSDGNSGAKTV 1080
Db 1021 DSSSDATSNANLTITKELKLTEDLSISGNKAEITAKOGRDLTIGNSDGNSGAKTV 1080
Qy 1081 TFNVKDSKISADGHNVTLNSKVTSSNGSGRESNDTGLTITAKNVEVNDKITSKLT 1140
Db 1081 TFNVKDSKISADGHNVTLNSKVTSSNGSGRESNDTGLTITAKNVEVNDKITSKLT 1140
Qy 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200
Db 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200
Qy 1201 KSGEANVTSATGTIGGTISGNTVNTANAGDLTVGNGAEINATGAATLTATGNTLTTEA 1260
Db 1201 KSGEANVTSATGTIGGTISGNTVNTANAGDLTVGNGAEINATGAATLTATGNTLTTEA 1260
Qy 1261 GSSITSTKGOVDLLAONGSIAGSINAANVTLLNTTGLTTTAVAGSDIKATSGTLVINAKDAK 1320
Db 1261 GSSITSTKGOVDLLAONGSIAGSINAANVTLLNTTGLTTTAVAGSDIKATSGTLVINAKDAK 1320
Qy 1321 LMGDASGDSSTEVNAVNASGSGVTAATSSSVNITGDLNTVNGLNTISKDGRNTVRLRGKE 1380
Db 1321 LMGDASGDSSTEVNAVNWGFGSVTAATSSSVNITGDLNTVNGLNTISKDGRNTVRLRGKE 1380
Qy 1381 IEVKYIQGVASVEVIEAKRVLEKVKDLSDEERETLAKGVSAVRPEPNNTITVANTON 1440
Db 1381 IEVKYIQGVASVEVIEAKRVLEKVKDLSDEERETLAKGVSAVRPEPNNTITVANTON 1440
Qy 1441 EFTTRPSSQVIIESEKACFSSGNGARVCTNVNADGQP 1477
Db 1441 EFTTRPSSQVIIESEKACFSSGNGARVCTNVNADGQP 1477

RESULT 6
AAW30292
ID AAW30292 standard; Protein: 1601 AA.
XX
AC AAW30292;
XX
DT 14-APR-1998 (first entry)
XX
DE Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
```

```
KW Non-typeable Haemophilus; high molecular weight surface protein;
XX HMW4; immunogen; vaccine; otitis media.
OS Haemophilus influenzae strain 5.
FH Key Location/Qualifiers
FT Misc-difference 372 /note= "encoded by TCT"
FT Misc-difference 400 /note= "encoded by AAT"
FT
XX WO9736914-A1.
PN
XX 09-OCT-1997.
PD
XX 01-APR-1997; 97WO-US04707.
PF
XX 01-APR-1996; 96US-0617697.
PR
XX (BARE/) BARENKAMP S J.
PA
XX Barenkamp SJ;
PI
XX
XX WPI: 1997-503038/46.
XX N-PSDB: AAT90993.
XX High molecular weight proteins of non-typeable Haemophilus
XX influenzae - useful for vaccine production
XX
XX Claim 1: Page 97-102; 183pp; English.
XX
XX This protein comprises the high molecular weight surface protein
XX HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
XX has the immunological ability to protect against disease caused by
XX a non-typeable Haemophilus strain and is characterised by at least
XX one surface-exposed B-cell epitope that is recognised by monoclonal
XX antibody A06. The HMW4 amino acid sequence was deduced from an
XX isolated HMW4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
XX AAW30294) and HMW3 (see AAW30291) have also been identified. A
XX conjugate comprising HMW4 linked to an antigen, hapten or
XX polysaccharide, and a synthetic peptide of 6-150 amino acids
XX corresponding to at least protective epitope of HMW4 are also
XX claimed. HMW proteins, conjugates and peptides can be used in
XX vaccines, as immunogens for preparation of antibodies and as
XX antigens for detection of these antibodies.
XX
XX Sequence 1601 AA:
XX
XX Query Match 71.3%; Score 5284; DB 18; Length 1601;
XX Best Local Similarity 68.4%; Pred. No. 3.4e-243;
XX Matches 1100; Conservative 136; Mismatches 231; Indels 142; Gaps 16;
Qy 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGEKSEKPARMKVRHLAKPLSALLSLGYT 60
Db 1 MNKIYRLKFSKRLNALVAVSELTRGCDHSTKGEKSEKPVRTKVRHLAKPLSALLSLGMA 60
Qy 61 SIPQSVLASGLQGMVDVVGHTATMVDGCKTIIIRNSVDAILNKKOFNIDQNMVQFLQENN 120
Db 61 SIPQSVLASGLQGMVDVVGHTATMVDGCKTIIIRNSVDAILNKKOFNIDQNMVQFLQESS 120
Qy 121 NSAVFNRYTSTNOTISQKGLDLSNGOVFLINPNGITITGKDAIINTNGTASTLDSINENIK 180
Db 121 NSAVFNRYTSTNOTISQKGLDLSNGOVFLINPNGITITGKDAIINTNGTASTLDSINENIK 180
Qy 181 ARNFTFQTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGIVSNGGSIISLAGOKIT 240
Db 181 ARNFTFQTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGIVSNGGSIISLAGOKIT 240
Qy 241 ISDIINPTITYSTAAPENEAVNLGDIKAGKGINVRAATIRNOCKLSADSVSKDGSNIV 300
Db 241 ISDIINPTITYSTAAPENEAVNLGDIKAGKGINVRAATIRNOCKLSADSVSKDGSNIV 300
Qy 301 LSAKEGEAEIGGVISAQNOQAQKGLMITGDKVTLTKTGAVIDLSGEGGEGTYLGGDERGE 360
XX
```

```

Db 301 LSAKEGEAEIGSVISAQNOQAKGKLMITGDKVTLTKTGAVIDUSGRGGETYLGGERGE 360
QY 361 GKGIOIAKKTSLBKSGSTINVSKEKGGFAIVMGDIALIDGNINAQSGGDIAGTKGFVET 420
Db 361 GKGIOIAKKTSLBKSGSTINVSKEKGGFAIVMGDIALIDGNINAQSG--DIAGTKGFVET 419
QY 421 SGHDLFTKDAIVDAKELWLDFOVNSINAEADPLFNWNTGDEPTGTGEASDPKKKSELK 480
Db 420 SGHDLSTGDDVVDVDAKELWLDFOVNSINAEADPLFNWNTGDEPTGTGEASDPKKKSELK 479
QY 481 TLTNTTISNLYKNAMTNTASRKLTVNSSINIGSNHLLHLSKSGORGQGGVQIDGDIITS 540
Db 480 PTLTNTSTLEQLLRGYSVYVNTANNRIVYVNSINL--SNGSLTLTKTGD--GVKINGDIITS 535
QY 541 -KGNLTIVSGWVDVHKNTITLDOGFELNTAA--SVAPEGGNKKARDAANAKIVAQGTVI 598
Db 536 NENGLTLFKAGSWVDVHKNTITLDOGFELNTAA--SVAPEGGNKKARDAANAKIVAQGTVI 595
QY 599 TGECKDFRANVSLNGTGKGLNLISSVNNLTINLSGTINISGNITINOTTRKNTSWQTS 658
Db 596 NKDDKQFRFNVSINGTGKGLNLISSVNNLTINLSGTINISGNITINOTTRKNTSWQTS 655
QY 659 HDSHWNVSALENGTGANFTFIKYI--SSNSKGLTTQYRSSAGVNFNGVNMSEFLKECA 716
Db 656 KDSYVWVSLTLNTVQKFTFIKFDVDSGNSQDLRSRRSFAGVHFGIGGKTNFNGANA 715
QY 717 KVFNLKPNENMNTSKPLRFLANITATGG--GSVFEDIYANHSGRGAELKMSINTSN 774
Db 716 KALFKLKPNAATDPKELPITFNANITATGNSDSSVMFEDIHANLTSRAAGINMDSINITG 775
QY 775 GANFTLVHVRGDDAFKINKOLTNATNSFNLSROTKDDFDGYARNAINSTYNTISILGG 834
Db 776 GLDFSITSHNRNSNAFEIKKOLTNATNSFNLSROTKDDFDGYARNAINSTYNTISILGG 835
QY 835 NVTLGSONSSSITGNITIEKAAVTLEANNAPNOQNTDRVYKIGSLVLLVNGSLTGEN 894
Db 836 NVTLGSONSSSITGNITIEKAAVTLEANNAPNOQNTDRVYKIGSLVLLVNGSLTGEN 895
QY 895 ADIKGNLTISEATPKGTRDTLTNITGNTNGTAEINITOGVVKL--GNVTNDGDLNITT 953
Db 896 ANIVGNLSIAEDSTFKGEASDNLNITGNTNGTAEINITOGVVKL--GNVTNDGDLNITT 955
QY 954 HAKRNORSIIGDDIINKKGLNITSDNDABEIQIGNITSQKEGNTLTISSDKINITKQITI 1013
Db 956 NASGTQKTLIINGNITNEKGLNITSDNDABEIQIGNITSQKEGNTLTISSDKINITKQITI 1015
QY 1014 KKGIDGEDSSDATSANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSDGNS 1073
Db 1016 KAGVEGGRSDSEANENLTIQTKELKLAGDLNLSGFNKAETAKDGRDLTIGNSDGNS 1074
QY 1074 GAETKVTFFNVKDSKISADGHNVTLSKVKTSNGGRESNDRGLTITAKNVEVVK 1133
Db 1075 -ADAKKVTDFKVKDSKISTDGHNTLNSEVKT--SNGSSNAGNDNSTGLTISAKDVTNN 1131
QY 1134 DITSLKTVNI-----TASEKVTI 1151
Db 1132 IVTSKHTINISAACNVTTEGTTINATTSVEVTAQNGTIKGNITISQNVTVTATENLVT 1191
QY 1152 TAGSTINATNGKASITTKTGDISG-----TISGNTVSVS-----ATVD--- 1189
Db 1192 TENAVINATSGTVNISYTKTGDKIGGISTSGNVTITAGSNLTKVSNITGODVTVTADAGA 1251
QY 1190 LTTKSGSKIEAKSGEANTVSATG----- 1212
Db 1252 LTTTATGTSIATTGNANITTKTGDKINGKVESSSGTVLTATGATLAVGNISGNTVITAD 1311
QY 1213 -----TIG-----GTISGNTVNTVANAGDLVGVNGAEINATEGAA 1247
Db 1312 SGKLTISVGTINGNTSVTSSSGDIEGTISGNTVNTVANAGDLVGVNGAEINATEGAA 1371
QY 1248 TLTATGNTLTTEAGSSITSTKGOVDLLAONGIAGSINAANVTNTTGLTTVVAGSDIKA 1307

```

```

Db 1372 TLTAESGKLTTOTGSSITSSNGQTTLTAKDSSIAGNINANVTNLTGTLTTTGSKINA 1431
QY 1308 TSGTLVINAKDAKLAGDASDSTEVNNAVNSGSGVTAATSSSVNITGDLNTVNGLNITS 1367
Db 1432 TSGTLTINAKDAKLAGDASDSTEVNNAVNSGSGVTAATSSSVNITGDLNTVNGLNITS 1491
QY 1368 KDCRNTVRLRGREIEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSARVF 1427
Db 1492 ENGRNTVRLRGREIEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSARVF 1551
QY 1428 VEPNNTITVNTQNEFTTRPSSQVITSEGKACFSSGNGARVCTNVAADDGQ 1476
Db 1552 VEPNNTITVNTQNEFTTRPSSQVITSEGKACFSSGNGARVCTNVAADDGQ 1600

```

RESULT 7

```

AAB01849
ID AAB01849 standard; Protein; 1036 AA.
AC
XX AAB01849;
XX
XX 11-SEP-2000 (first entry)
XX
XX Haemophilus influenzae strain 12 mature HMW2A protein, SEQ ID NO:73.
DE
XX Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
XX Haemophilus influenzae strain 12.
OS
XX WO200020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WO-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-303789/26.
XX
XX N-PSDB; AAA52198.
XX

```

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

Claim 8; Fig 29E-N; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the 77 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMW proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMW. The invention also discloses hmwA genes (AAA52175-A52198) and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Joysc. K1, K21, LDC22, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in

CC humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HmW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, CC HmW proteins and/or HmW peptides. The nucleotide sequences encoding the CC HmW proteins can be used to isolate and clone hmw genes from other CC non-typeable strains of Haemophilus via hybridisation reactions. The CC -present sequence represents a mature HmW protein from a non-typeable CC strain of H. influenzae.

XX
SQ Sequence 1036 AA;

Query Match 69.7%; Score 5164; DB 21; Length 1036;
Best Local Similarity 99.5%; Pred. No. 1e-237;
Matches 1030; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 443 DNVSNIAEDPLENNTGINDPEFTGTGEADPKKNSLKTTLTNTTISNVLKNAWMTNITA 502
Db 2 DDVTEADPLRNTGINDPEFTGTGEADPKKNSLKTTLTNTTISNVLKNAWMTNITA 61
Qy 503 SRKLTVNSSINIGNSHLILHSGGORGGOVDGDIITSGKGLTIYSGGWVDVHKNTILD 562
Db 62 SRKLTVNSSINIGNSHLILHSGGORGGOVDGDIITSGKGLTIYSGGWVDVHKNTILD 121
Qy 563 OGFLNITAAVAPEGGNKKARDANAKIVAQGVTTITGEGKDFRANNSVLNGTGKGLNII 622
Db 122 OGFLNITAAVAPEGGNKKARDANAKIVAQGVTTITGEGKDFRANNSVLNGTGKGLNII 181
Qy 623 SSVNLTNLSGTTINISGNTITINQTRKNTSYQWTSKSHWNSALNLETGANFTFIKYI 682
Db 182 SSVNLTNLSGTTINISGNTITINQTRKNTSYQWTSKSHWNSALNLETGANFTFIKYI 241
Qy 683 SSNSKGLTQYRSSAGVNFNGVNGNMSFNLEKAGVNFKLKPNENMNTSKPLPIRELANI 742
Db 242 SSNSKGLTQYRSSAGVNFNGVNGNMSFNLEKAGVNFKLKPNENMNTSKPLPIRELANI 301
Qy 743 TATGGGVFFDIYANHSGRGAELKMSINISNGANFTLNSHVRGDDAFKINKDLTINATN 802
Db 302 TATGGGVFFDIYANHSGRGAELKMSINISNGANFTLNSHVRGDDAFKINKDLTINATN 361
Qy 803 SNFSLROTQDFYDGYARNAINSTYNIISILGGNVTLGGQNSSSSITGNITIEKAANVTLE 862
Db 362 SNFSLROTQDFYDGYARNAINSTYNIISILGGNVTLGGQNSSSSITGNITIEKAANVTLE 421
Qy 863 ANNAPNOQNIIRDYIKLGLSLVNGSLTGTENADIKGNLTISESATPKGKTRODTLNTGN 922
Db 422 ANNAPNOQNIIRDYIKLGLSLVNGSLTGTENADIKGNLTISESATPKGKTRODTLNTGN 481
Qy 923 FTNNGTAEINTQGVVKGNGVNTDGLNITTHAKRNORSIIGGDIINKKGLSLNTDSND 982
Db 482 FTNNGTAEINTQGVVKGNGVNTDGLNITTHAKRNORSIIGGDIINKKGLSLNTDSND 541
Qy 983 AEIQIGGNISOKGNLTISDKINITYITIKKIGIDGESSDATSANLTITKTELKLT 1042
Db 542 AEIQIGGNISOKGNLTISDKINITYITIKKIGIDGESSDATSANLTITKTELKLT 601
Qy 1043 EDLSISGFNKAETAKDGRDLTIGNSDNGNSGAEAKVTTFNNVKDSKISADGHNVTLSK 1102
Db 602 EDLSISGFNKAETAKDGRDLTIGNSDNGNSGAEAKVTTFNNVKDSKISADGHNVTLSK 661
Qy 1103 VKTSSNGGRESNDNDTGLTITARNVENVNKDITSLKTVNITASEKVTITAGSTINATNG 1162
Db 662 VKTSSNGGRESNDNDTGLTITARNVENVNKDITSLKTVNITASEKVTITAGSTINATNG 721
Qy 1163 KASITTKTGDISGTSIGNTVSVATVDTLTTKSGSKIEAKSGEANTVATGCTIGCTISGNT 1222
Db 722 KASITTKTGDISGTSIGNTVSVATGDLTTKSGSKIEAKSGEANTVATGCTIGCTISGNT 781
Qy 1223 VNVATANAGDLVNGCAENATNTEGATLTATGNTLTTEAGSSITTSKQVDLLAONGSIAG 1282
Db 782 VNVATANAGDLVNGCAENATNTEGATLTATGNTLTTEAGSSITTSKQVDLLAONGSIAG 841
Qy 1283 SINAANVTLTNTGTLTVAGSDIKATSGTLVINAKDAKINGDASGDSTEVENAVNASGSGS 1342
Db 1283 SINAANVTLTNTGTLTVAGSDIKATSGTLVINAKDAKINGDASGDSTEVENAVNASGSGS 1021

Db 842 SINAANVTLTNTGTLTVAGSDIKATSGTLVINAKDAKINGDASGDSTEVENAVNASGSGS 901
Qy 1343 VTAATSSSVNITGDLNTVNGLNIIISKDGRNIVRLRGKEIEVKYIOPGVASVEEVEAKRV 1402
Db 902 VTAATSSSVNITGDLNTVNGLNIIISKDGRNIVRLRGKEIEVKYIOPGVASVEEVEAKRV 961
Qy 1403 LEKVKDLSDERETLAKLGVSARVRENNITVNTONEFTTRPSSQVILISEGKACFSSG 1462
Db 962 LEKVKDLSDERETLAKLGVSARVRENNITVNTONEFTTRPSSQVILISEGKACFSSG 1021
Qy 1463 NGARVCTNVADGQP 1477
Db 1022 NGARVCTNVADGQP 1036

RESULT 8
AAR41723
ID AAR41723 standard; Protein; 1536 AA.
XX
AC AAR41723;
XX
DT 26-APR-1994 (first entry)
XX
DE High molecular weight protein 1 (HmW1).
XX
DE HmW: high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX
OS Haemophilus Influenzae.
XX
PN W09319090-A.
XX
PD 30-SEP-1993.
XX
PF 16-MAR-1993; 93WO-US02166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
PA (BARE/) BARENKAMP S J.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Barenkamp SJ;
XX
WPI: 1993-320683/40.
DR N-PSDB: AAQ49506.
XX
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
PS Claim 3; Figure 2; 100pp; English.
XX
CC The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX
SQ Sequence 1536 AA;

Query Match 67.1%; Score 4967; DB 14; Length 1536;
Best Local Similarity 67.3%; Pred. No. 3.9e-228;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKESKPARMKVRHLAKPLSAMLISLGV 60
Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKESKPARMKVRHLAKPLSAMLISLGV 60
Qy 61 SIFQSVLASGLQGMVVGHTATMOVGDKTIIRNSDAIINWKQFNIDQNMVQFLOENN 120
Db 61 SIFQSVLASGLQGMVVGHTATMOVGDKTIIRNSDAIINWKQFNIDQNMVQFLOENN 120
Qy 121 NSAVFNKRVTSNQISQLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDISNENIK 180
Db 121 NSAVFNKRVTSNQISQLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDISNENIK 180

Db 121 NSAVENRVTSNQISOLKGLDSNGQVFLINPNPNIITIGKDAIINTNGFTASTLDISNENIK 180
QY 181 ARNFFTEQTKKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSISSLAGQKIT 240
Db 181 ARNFFTEQTKKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSISSLAGQKIT 240
QY 241 ISDIINPTIYSIAAPENAEVNLGDFIPAKGGNNINRAATIRNOGKLSADSVSKDKNIV 300
Db 241 ISDIINPTIYSIAAPENAEVNLGDFIPAKGGNNINRAATIRNOGKLSADSVSKDKNIV 300
QY 301 LSAKEGEAEIGVISAQOQAGKGLMTITGDKVTLTKTCAVIDLGSKEGGEYVLGGDERGE 360
Db 301 LSAKEGEAEIGVISAQOQAGKGLMTITGDKVTLTKTCAVIDLGSKEGGEYVLGGDERGE 360
QY 361 KGNGIOLAKTISLEKGSTINYSKGEKGFALVWGDIALIDGNNINAQSGDIAKTGGFVET 420
Db 361 KGNGIOLAKTISLEKGSTINYSKGEKGFALVWGDIALIDGNNINAQSGDIAKTGGFVET 420
QY 421 SGHDLFIKDNAIDVAKEWLLDFDWNVINAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479
Db 421 SGHDLFIKDNAIDVAKEWLLDFDWNVINAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479
QY 479 KTTLTNTTLESILKGFVNITANQRIYVNSSINL-SNGSLTLWSEGRSGGVGVEINNDIT 537
Db 479 KTTLTNTTLESILKGFVNITANQRIYVNSSINL-SNGSLTLWSEGRSGGVGVEINNDIT 537
QY 537 TGDTRGANLTIYSGGWVDEHKNISLGAQGNINITAKQDIAFEKGSQV-----ITGQ 590
Db 537 TGDTRGANLTIYSGGWVDEHKNISLGAQGNINITAKQDIAFEKGSQV-----ITGQ 590
QY 594 GTVITITGEGKDFRANNVSLNCTGKGLNISSVNN--LTHNLSCVINTSGNITNQTRK 650
Db 594 GTVITITGEGKDFRANNVSLNCTGKGLNISSVNN--LTHNLSCVINTSGNITNQTRK 650
QY 650 NTS-YWOTSHDHNWVSALETCANFTF-IKYISSSKGLTTQYRSSAGVNFNGVANGNM 708
Db 650 NTS-YWOTSHDHNWVSALETCANFTF-IKYISSSKGLTTQYRSSAGVNFNGVANGNM 708
QY 708 NESGYDFKFGRTYNNLTSVNESGEENLTIDSRGSDSAGTLTQYNLNGISF--NKDT 706
Db 708 NESGYDFKFGRTYNNLTSVNESGEENLTIDSRGSDSAGTLTQYNLNGISF--NKDT 706
QY 764 SFNLKEGAKVNFKLKPNENMTSKPLP-IRFLANITATGGSGVFFDIYANHSU---RGAE 764
Db 764 SFNLKEGAKVNFKLKPNENMTSKPLP-IRFLANITATGGSGVFFDIYANHSU---RGAE 764
QY 766 TFNVERNARVNDIKAPIGIKYSSSLVAFNGNISVSGGSDVFTLLASSNNVQTECV 766
Db 766 TFNVERNARVNDIKAPIGIKYSSSLVAFNGNISVSGGSDVFTLLASSNNVQTECV 766
QY 824 LKMSINISGANFTLNSHVGRDADFINKDLTINATNSNFSLRQTKDDFDYGVARNAIN 824
Db 824 LKMSINISGANFTLNSHVGRDADFINKDLTINATNSNFSLRQTKDDFDYGVARNAIN 824
QY 824 INSKYFNVSTGSSLRFTSGTKTGFSTERDLTLNATGNNITLLOVEGT--DGMIGKIV 824
Db 824 INSKYFNVSTGSSLRFTSGTKTGFSTERDLTLNATGNNITLLOVEGT--DGMIGKIV 824
QY 884 STYNISLGGNVTLGGONSSSTGNTTIEKAAVNTLEANNAPQOIRDRVILKGLSLV 884
Db 884 STYNISLGGNVTLGGONSSSTGNTTIEKAAVNTLEANNAPQOIRDRVILKGLSLV 884
QY 882 AKKNITTEGGNITFGSRKAVTEIEGNTINNANVTILIGSDFDNHQ--KPLTIKKDVIIN 882
Db 882 AKKNITTEGGNITFGSRKAVTEIEGNTINNANVTILIGSDFDNHQ--KPLTIKKDVIIN 882
QY 944 NGSLSLTGENADIKGLNLTISESATEFKGTRDTLNTGNTNFTNGTAETINITQGVVKGNGVT 944
Db 944 NGSLSLTGENADIKGLNLTISESATEFKGTRDTLNTGNTNFTNGTAETINITQGVVKGNGVT 944
QY 942 SGNLTAGNVINAGNLTVESNFKAITNFTFNVGGLFDKNGNSNISIAKGGARFKDID 942
Db 942 SGNLTAGNVINAGNLTVESNFKAITNFTFNVGGLFDKNGNSNISIAKGGARFKDID 942
QY 1004 NQDLNITTHAKNRQSIIGDIIKNGSLNITDSNDAETIQIGNITSQEGNLTISDDK 1004
Db 1004 NQDLNITTHAKNRQSIIGDIIKNGSLNITDSNDAETIQIGNITSQEGNLTISDDK 1004
QY 1002 NSKNLSITSSSYRPIISGNTNKNKNDLNIINEGSDTEWQIGDVYSQEGNLTISDDK 1002
Db 1002 NSKNLSITSSSYRPIISGNTNKNKNDLNIINEGSDTEWQIGDVYSQEGNLTISDDK 1002
QY 1064 INITKQITIKKIDGEDSSDATSNANLTIKTRELKLTEDLSISGPNKAEITAKDGRDLT 1064
Db 1064 INITKQITIKKIDGEDSSDATSNANLTIKTRELKLTEDLSISGPNKAEITAKDGRDLT 1064
QY 1062 INITKQITIKAGVDGENSDATNANLTIKTRELKLTQDLNISGPNKAEITAKDGSDLT 1062
Db 1062 INITKQITIKAGVDGENSDATNANLTIKTRELKLTQDLNISGPNKAEITAKDGSDLT 1062
QY 1124 IGSNDGNSCAEAKVTFFNVKDKISADGHNTVLSKVTSSNGRGRESNDTGLTI 1124
Db 1124 IGSNDGNSCAEAKVTFFNVKDKISADGHNTVLSKVTSSNGRGRESNDTGLTI 1124
QY 1121 IGNTNSAD-GTNNAKVTFFNVKDKISADGHNTVLSKVTSSNGRNNNTDSSDNNAGLTI 1121
Db 1121 IGNTNSAD-GTNNAKVTFFNVKDKISADGHNTVLSKVTSSNGRNNNTDSSDNNAGLTI 1121
QY 1170 TAKNVEYKNDITSLKTVNITA-SEKVTITTAGSTINATNGKASITTKT----- 1170
Db 1170 TAKNVEYKNDITSLKTVNITA-SEKVTITTAGSTINATNGKASITTKT----- 1170
QY 1181 DAKNVTNNITSHKAVISISATSEITTKGTINATNGVETIAQTGSLGIESSGS 1181
Db 1181 DAKNVTNNITSHKAVISISATSEITTKGTINATNGVETIAQTGSLGIESSGS 1181
QY 1181 -----GDISGTISGNT 1181
Db 1181 -----GDISGTISGNT 1181
QY 1241 VTLTATEGALAVSNIISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDIGTISGGT 1241
Db 1241 VTLTATEGALAVSNIISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDIGTISGGT 1241

QY 1182 VSVSATVDLTTKSGKIEAKSGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1241
Db 1242 VEVKATESLTTOSNSKIKATGTGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1301
QY 1242 ATEGAATLTATGNTLTTEAGSSITSTKGQVLLLAQNGSIAGSINAANVTLTNTTGLTTVA 1301
Db 1302 ATEGAATLTSSGKLTTEASSHITSAGQVNLQAQSGSVAGSINAANVTLTNTTGLTTVK 1361
QY 1302 GSDIKATSGTLVINAKDAKLNGSDSTEVNAVNASGSGSVTAATSSVNTTGLDNTVN 1361
Db 1362 GSNINATSGTLVINAKDAELNGALGNHTVVNATNANGSGSVIATTSRVNTTGLDITIN 1421
QY 1362 GLNIISKNGINTVLLKGVKIDVYIQPGIASVDEIEAKRILEKVKDLSDEERETLAKLG 1421
Db 1422 GLNIISKNGINTVLLKGVKIDVYIQPGIASVDEIEAKRILEKVKDLSDEERETLAKLG 1481
QY 1482 VSAVREVPNNITVNTQNEFTTREPSSQVIISSEKACFSSNGARVCTNVADDDQ 1476
Db 1482 VSAVREVPNNITVNTQNEFATPLSRIVISEGRACFSNSDGTATVCVNIADNGR 1536

RESULT 9

AAR41732

ID AAR41732 standard; Protein; 1529 AA.

XX AAR41732;

XX AC
XX 26-APR-1994 (first entry)

XX High molecular weight protein 4 (HMW4).

XX HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.

XX Haemophilus influenzae.

XX PN W09319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993; 93WO-US02166.

XX 16-MAR-1992; 92GB-0005704.

XX (BARE/) BARENKAMP S J.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.
XX N-PSDB; AAQ49511.XX High molecular weight surface proteins - of non-typeable
XX haemophilus which exhibit immunogenic properties

XX Claim 6; Figure 10; 100pp; English.

XX The isolation and purification of the high molecular weight protein
XX enables the identification of the major protective epitopes of the
XX protein by conventional epitope mapping. These epitopes can then be
XX synthesised using standard techniques and incorporated into fully
XX synthetic or recombinant vaccines.

XX Sequence 1529 AA;

Query Match 66.9%; Score 4957; DB 14; Length 1529;
Best Local Similarity 67.2%; Pred. No. 1.2e-227;

Matches 1033; Conservative 195; Mismatches 227; Indels 142; Gaps 16;

QY 73 GMDVHGTATMQVDGKNTIIRNSVDALINNKQFNIDQNEVQFLOENNSAVFNRTSNQ 132
Db 1 GMSVHGTATMQVDGKNTIIRNSVDALINNKQFNIDQNEVQFLOENNSAVFNRTSDQ 60

QY 133 ISOLKGLDSNGVFLINPGITIGKDAIINTNGFTASTLIDISNENIKARNFTFEOTKDK 192
 Db 61 ISOLKGLDSNGVFLINPGITIGKDAIINTNGFTASTLIDISNENIKARNFTLEOTKDK 120
 QY 193 ALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSSLLAGOKITISDIINPIITS 252
 Db 121 ALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSSLLAGOKITISDIINPIITS 180
 QY 253 IAPENEAVALGDIIPAKGGNINRAATIRNOGKLSADSVSKSGNIVLSAKEGEAIEGG 312
 Db 181 IAPENEAVALGDIIPAKGGNINRAATIRNOGKLSADSVSKSGNIVLSAKEGEAIEGG 240
 QY 313 VISAQNOQAKGGKMITGDKVTLTKTGAVIDLSGKEGETYLGDERGEGKNGIOLAKKTS 372
 Db 241 VISAQNOQAKGGKMITGDKVTLTKTGAVIDLSGKEGETYLGDERGEGKNGIOLAKKT 300
 QY 373 LEKSTINVSKEKGGFAIVMGDIALIDGNINAQGGSDIAKTGGFVETSGHDLFIKDNAI 432
 Db 301 LEKSTINVSKEKGGFAIVMGDIALIDGNINAQGS-DIAKTGGFVETSGHDLISIGDDVI 359
 QY 433 VDAREWLLDPNVISNAEDPLENNTGIDNFTGTGEASDPKKNSELKTLTNTTISNYL 492
 Db 360 VDAREWLLDPNVISNLTSGRNNNGENQGYTTGDKTKESPKGNSLSKPLTLNSTLEQIL 419
 QY 493 KNAFTMITASRKLTVNSINSGNSHLILHSKGORGGOVIGDIDITS-KGGLNLTIVSGG 551
 Db 420 RRGSVNITANRIYVNSSINL-SNGSLTLTKRD---GVKINGDITSNENGLTIKAGS 475
 QY 552 WDVHKNITLDOGLNITAA-SVAFEGGNKARDAANAKIVAOGTVITITGEGKDFRANV 610
 Db 476 WDVHKNITLDOGLNITVAGSVAFEGGKARNATDAQITAOGTITVFNKDDKQFRFNV 535
 QY 611 SLNGTGKGLNITISSVNLTHLSGTINISGNITITQTRKNYSWQTSBHSNVSALNL 670
 Db 536 SINGTGKGLKFIANONFTHKFDEGEINISGIVTINQTKDKVKNYNAKSDSYNVSLSLT 595
 QY 671 ETGANFTPIKYI--SSSKGLITTOYRSSAGVNFNGVNGNMSFNLKEGAKYFNKLPKNEM 728
 Db 596 NTVOKFTPIKFDVSGNSQDLRSRRSFAGVHFNGIGGKTNFNIGANAKALFKLPNAAT 655
 QY 729 NTSKPLPIRFLANITATGG--GSVFEDIYAHNSGRGELKMSSEINISGANAMFTLSNVRG 786
 Db 656 DPKKELPITFNANITATGSDSSVMEDIHANLTSRAAGINMDSINITGGLDIFSITSHRN 715
 QY 787 DDAKINKDLTINATNSFSLROTCKDDPYDGYARNAINSTYNIISILGNYTLGGONSSS 846
 Db 716 SNAFEIKDLTINATNSFSLKOTKDSFYNEYSKHAINSHNLILGNYTLGGONSSS 775
 QY 847 ITGNITIEKAANVLEANNAPNOQNIORRVTKLGSLLVNGSLTGENADIKNLTISES 906
 Db 776 ITGNITIEKANVTLOADTSNGLKRTLTIGNISVEGSLTGANANIVGNLSIAED 835
 QY 907 ATFKGKTRDTLNTGNFTNCTAEINITQGVVKL-GNVTNDGDLNITHAKRNORSIIG 965
 Db 836 STFKGEASDNLNITFTFNGTANINIKQGVVKLQGDILINKGGLNITNASGTOKTIING 895
 QY 966 DIINKKGLNITDSNNDIAEQIGGNISQKEGNLTISSDKINITKQITKKGIDGEDSSD 1025
 Db 896 NITNEKGLNITKNIKADAEIQIGGNISQKEGNLTISSDKVNITQITKAGVEGGRSDSS 955
 QY 1026 ATSNANLTITKTELKLTEDLSGTSGFNKAPITAKGRDITLIGNSDNGSGAEAKTVTFNV 1085
 Db 956 EAENANLTITKTELKLAGDLNISGFNKAELITAKNGSDLTIGNASGGN--ADAKKVTFDK 1013
 QY 1086 KDSKISADGHNVTLNSKVKTSNGSGRNSDNDTGLTITAKNVEYVKNKDTISLAKTVAI-- 1143
 Db 1014 KDSKISTDGHNVTLNSKVKTSNGSGRNSDNDTGLTITAKNVEYVKNKDTISLAKTVAI-- 1071
 QY 1144 -----TASEKVTITAGSTINATNGK 1163
 Db 1072 AAGNVTTKEGTTINATTSVEVTAQNGTIKGNITSONVTVTATENLVTENAVINATSGT 1131

QY 1164 ASITTKTGDISG-----TISGNTSVSVS-----ATVD---LTTKSGSKIEAK 1201
 Db 1132 VNISTKTDIKGIESTSGNVNITASGNTLKVSNITGQDVTVTADAGALTITAGSTISAT 1191
 QY 1202 SGEANVTSATG-----TIG--- 1215
 Db 1192 TGNANITTKTGDISGCKVRESSGSVTLVATGATLVAGNISGNITVITADSGKLTSTVGSTI 1251
 QY 1216 -----CTISGNTVNTANAGDLTVNGCAEINATGCAATLTATGNTLTTE 1259
 Db 1252 NGTNSVTTSSQSDIEGTSIGNVTNVYATSTGDLTIGNSKAVEKNGAATLTAESGKLTQ 1311
 QY 1260 AGSSTSTKQVDVLLAONGSIAGSINAANVTLLTNTTGLTTVAGSDIKATSGTLVINAKDA 1319
 Db 1312 TGSSTISSNGQTLTAKDSSIAAGNINAANVTLLTNTTGLTTGDSKINATSGTLTINAKDA 1371
 QY 1320 KLNGDASGSTEVNAVNASGSGSVTAATSSSVNITGDLTVNGLNIISKDGRNVTVLRGK 1379
 Db 1372 KLDGAASGDRVTYVNTATNAGSGNVTAKTSSSVNITGDLTVNGLNIISNGRNTVLRGK 1431
 QY 1380 EIEVKYIQGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPVPEPNTITVNTQ 1439
 Db 1432 EIDVKYIQGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPVPEPNTITVNTQ 1491
 QY 1440 NEFTTRPSQVITISEGKACFSSGNGARVCTNVADDGQ 1476
 Db 1492 NEFTTRPSQVITISEGKACFSSGNGARVCTNVADDGQ 1528

RESULT 10

AAW30293

ID AAW30293 standard; Protein; 1536 AA.

XX AAW30293;

DT 14-APR-1998 (first entry)

DE Non-typeable Haemophilus high mol.wt. surface protein HMW1.

KW Non-typeable Haemophilus; high molecular weight surface protein; HMW1; hmw1A gene; immunogen; vaccine; otitis media.

XX Haemophilus influenzae strain 12.

XX Key Location/Qualifiers

FT Misc-difference 4 /note= "encoded by CTA"

FT Misc-difference 98 /note= "encoded by GAT"

FT Misc-difference 363 /note= "encoded by AAG"

XX W09736914-AL.

XX 09-OCT-1997.

XX 01-APR-1997; 97WO-US04707.

XX 01-APR-1996; 96US-0617697.

XX (BARE/) BARENKAMP S J.

XX Barenkamp SJ;

XX N-PSDB; AAT90994 and AAT90996.

XX High molecular weight proteins of non-typeable Haemophilus influenzae - useful for vaccine production

XX Claim 7; Page 66-70; 183pp; English.

XX This protein comprises the high molecular weight surface protein

CC

CC HMW1 (125 kDa) of non-typeable Haemophilus influenzae strain 12 that
 CC has the immunological ability to protect against disease caused by a
 CC non-typeable Haemophilus strain and is characterised by at least
 CC one surface-exposed B-cell epitope that is recognised by monoclonal
 CC antibody Ab6. The HMW1 amino acid sequence was deduced from the
 CC hmw1 gene sequence (see AAT90994 and AAT90996). The expressed protein
 CC is truncated, starting at residue 442 of the full-length gene
 CC product. HMW2 (see AAW30294), HMW3 (see AAW30291) and HMW4 (see
 CC AAW30292) have also been identified. A conjugate comprising HMW1
 CC linked to an antigen, hapten or polysaccharide, and a synthetic
 CC peptide of 6-150 amino acids corresponding to at least protective
 CC epitope of HMW1 are also claimed. HMW proteins, conjugates and
 CC peptides can be used in vaccines, as immunogens for preparation of
 CC antibodies and as antigens for detection of these antibodies.
 XX
 SQ Sequence 1536 AA;

Query Match 66.9%; Score 4956; DB 18; Length 1536;
 Best Local Similarity 67.3%; Pred. No. 1.3e-227;
 Matches 1046; Conservative 146; Mismatches 265; Indels 98; Gaps 20;

QY 1 MNKIVLKFESKRLNALVAVSELARGCDHSTEGSEKPARMKVRHLAKPLSAMLISLGV 60
 Db 1 MNKIVLKFESKRLNALVAVSELARGCDHSTEGSEKPARMKVRHLAKPLSAMLISLGV 60
 QY 61 SIQSVLASGLQMDVYHGTATMQVDGKNTIIRNSVDALINQKFNIDONEMVQFLQNN 120
 Db 61 SIQSVLASGLQMDVYHGTATMQVDGKNTIIRNSVDALINQKFNIDONEMVQFLQNN 120
 QY 121 NSAVFNRTVSNISQLKGLDSNGQVFLINPNTIGKDAIINTNGFTASTLDSINENIK 180
 Db 121 NSAVFNRTVSNISQLKGLDSNGQVFLINPNTIGKDAIINTNGFTASTLDSINENIK 180
 QY 181 ARNFTFEQTKKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSISLLAQKIT 240
 Db 181 ARNFTFEQTKKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSISLLAQKIT 240
 QY 241 ISDIINPTITYSIAAPEAVNLGDFIAKGGNINRAATIRNQGKLSADSVSKDKSGNIV 300
 Db 241 ISDIINPTITYSIAAPEAVNLGDFIAKGGNINRAATIRNQGKLSADSVSKDKSGNIV 300
 QY 301 LSAKEGEAEIGVISAQNOQAGKGLMITGDKVTLTKTCAVIDLSKEGEGTYLGGDERGE 360
 Db 301 LSAKEGEAEIGVISAQNOQAGKGLMITGDKVTLTKTCAVIDLSKEGEGTYLGGDERGE 360
 QY 361 GKNGIOLAKKTSLEKGSTINVSKEGEGFAIWWGDIALIDGNNIAQSGDIAKTGGFVET 420
 Db 361 GKNGIOLAKKTSLEKGSTINVSKEGEGFAIWWGDIALIDGNNIAQSGDIAKTGGFVET 420
 QY 421 SGHDLFIKDNAIVDAKEMLLDFDNVSNIAEDPLFNNGTINDEFPTGTG-EASDPKKNSEL 479
 Db 421 SGHDLFIKDNAIVDAKEMLLDFDNVSNIAEDPLFNNGTINDEFPTGTG-EASDPKKNSEL 479
 QY 480 KTTLTNTTISNKLNAWMTMNTTASRKLTVNSINIGSNHLLHSKGGGQVQIDGIT 539
 Db 479 KTTLTNTTISNKLNAWMTMNTTASRKLTVNSINIGSNHLLHSKGGGQVQIDGIT 539
 QY 540 ----SKGNLTISGGWVDVHKNTLD-QGFLNITA-ASVAFEGGNKKARDAANAIVAQ 593
 Db 538 TGGDTRGANLTISGGWVDVHKNTLSLAQAGNINITAKODIAFEKGSQV-----ITGQ 590
 QY 594 GTVTITGEGKDFRANNVSLNLTGKGLNIISSVNN---LTHNLSTINISGNITINOTTRK 650
 Db 591 GTIT-SGNQKGFNFNNVSLNLTGSGLOFTTKRTNKYAITNKEFTLNTSGKVNISMVLPK 649
 QY 651 NTS-YWTSHTSDSHWNVALNETCANFTF-IKYISSKSGLTQYRSAGVNFVNGVNGNM 708
 Db 650 NESGYDFKGTFTYNNLTSLNVSSEGFENLIDSRGDSAGTLTQPYNLNGISF---NKDT 706
 QY 709 SFNLKEGAKVNFKLKPNENMNTSKPLP-IRFLANITATGGSGVFPFIYANHSG---RGAE 764
 Db 707 TFNVARNARVNFIDIKAPIGINKYSSSLNVSFNGNISVSGGSDFTLLASSSNVQTPGVV 766

QY 765 LKMEINISGANFTLNSHVRGDDAFKINKDLTINATNSNFSLRQTKDDFDYGVARNAIN 824
 Db 767 INSKYFNVSTGSSLRFKTSGTSGTGFSEKDLTLNATGNNITLLQVEGT--DGMIGKIV 824
 QY 825 STYNISILGONVTLGCGNSSSTTGNITIEKAAVNTLEANNAPNOQINRDIRVILGSLV 884
 Db 825 AKKNITFEGNITFGSRKAVTEIEGNVTINNANVTILGSDFDNHQ--KPLTIKKDVIIN 882
 QY 885 NGSLSLTGENADIKGNLTISESATKCTRTDLNITGNFTNNGTAENITQGVVKGNTV 944
 Db 883 SGNLTAGGNIVNLAGNLTVESNANFAITNFTNVGGLFDNKGNSNISIAKGAARFDKD 942
 QY 945 NQDGLNITTHAKRNQRSIIIGDIIINKKGLNITDSNNDAEIQIGNISQKGNLTISDDK 1004
 Db 943 NSKNLSITTNSSSTYRTIISGNITNKNGLNITNEGSDTEWQIGGVSOKEGNLTISDDK 1002
 QY 1005 INITKOITIKKIGDGEDSSDSDATSNANLTIKTTELKLTEDLSISGFNKAEITAKDGRDLT 1064
 Db 1003 INITKOITIKAGVDGSDSDATSNANLTIKTTELKLTEDLSISGFNKAEITAKDGRDLT 1062
 QY 1065 IGNSDNGSNGAEAKVTFTFNNVKDSKISADGHNVTLASKVKYKTSNGSGRESNDTGLTI 1124
 Db 1063 IGMTNSAD-GTNAKKVTFNQVKDSKISADGHKVLKSHKVTSGSNNTHTSHSSDNNAGLTI 1121
 QY 1125 TAKNVEVNDKIDTSLKTVNITA-SEKVTITTAGSTINATNGKASITTKT----- 1170
 Db 1122 DAKNVTNNNITSHKAVSISATSGEITTKTGTINATGNVEITAGTSLGIESSSGS 1181
 QY 1171 -----GDISGTTISGNT 1181
 Db 1182 VLTATEGALAVNSISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDIGTTISGGT 1241
 QY 1182 VSVSATVDLTTSKSGKIEAKSGEANTVTSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1241
 Db 1242 VEVKATESLTQTSNKSIAKATTGEANTVTSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1301
 QY 1242 ATEGAATLTATGNTLTTEAGSITSTKQVDDLLAQNQSTAGSINAAVNTLTGTTTVA 1301
 Db 1302 ATEGAATLTSSGKLTTEASSHITSAGQVNLISAQDGSVAGSINAAVNTLTGTTTVA 1361
 QY 1302 GSDIKATSTLVINAKDADKINGDASGDSPEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1361
 Db 1362 GSNINATSTLVINAKDAELNCAALGNHTVNAVATNANGSGSVIATSSRVNITGDLNTVN 1421
 QY 1362 GLNIIISKGRNTRVRGKEIEVKYIQPGVASVEVEIEAKRVLEKVKDLSDEERETLAKG 1421
 Db 1422 GLNIIISKGRNTRVRGKEIEVKYIQPGVASVEVEIEAKRVLEKVKDLSDEERETLAKG 1481
 QY 1422 VSAVFEVPPNNTITVNTONEFTTREPSSOVIISEGRACFSSGNGCARVCTNVAADGQ 1476
 Db 1482 VSAVFEVPPNNTITVNTONEFTTREPSSOVIISEGRACFSSGNGCARVCTNVAADGQ 1536

RESULT 11
 AAR63505
 ID AAR63505 standard; Protein; 1536 AA.
 XX AAR63505;

AC AAR63505;
 XX 25-JUN-1995 (first entry)

DT Haemophilus high molecular weight protein HMW1.

XX High molecular weight protein; HMW1; protective vaccine; otitis;
 KW sinusitis; bronchitis; Hib.

OS Haemophilus.

PN WO9421290-A.

XX 29-SEP-1994.

PD 15-MAR-1994; 94WO-US02550.

XX 16-MAR-1993; 93US-0038682.
 XX (BARE/) BARENKAMP S J.
 XX (SGEM/) ST GEME J W.
 XX Barenkamp SJ, St GEME JW;
 XX WPI; 1994-316665/39.
 XX Q-PSDB; Q72293.
 XX New immunogenic high mol. wt. proteins of non typeable
 XX Haemophilus - useful in protective vaccines
 XX Claim 2; Page 31; 127pp; English.
 XX The HmW1 protein encoded by this sequence is useful in a vaccine to
 XX protect against disease caused by non-typeable Haemophilus which are
 XX not controlled by H. influenzae type b (Hib) vaccines. The encoded
 XX protein can also be used as a carrier for protective Hib
 XX polysaccharide (in a conjugate vaccine against meningitis) or for
 XX other antigens, haptens, etc.
 XX
 XX Sequence 1536 AA:
 Query Match 66.7%; Score 4943; DB 15; Length 1536;
 Best Local Similarity 67.1%; Pred. No. 5.4e-227;
 Matches 1043; Conservative 147; Mismatches 267; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKGESEKPARMKVRHLALKPLSALLSLGVT 60
 Db 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKGESEKPARMKVRHLALKPLSALLSLGVT 60
 Qy 61 SIPOSVLASGQGMVHVGTATMVDGDKTIIRNSVDIIINWKQFNIDQEMVQFLOENN 120
 Db 61 SIPOSVLASGQGMVHVGTATMVDGDKTIIRNSVDIIINWKQFNIDQEMVQFLOENN 120
 Qy 121 NSAVFNRTVSNQISQLKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDSNENIK 180
 Db 121 NSAVFNRTVSNQISQLKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDSNENIK 180
 Qy 181 ARNTEFQTKDKALAEIVNHLGIIIVGKDGSVNLIGKVGKNEGYSVNGGSIISLLAGOKIT 240
 Db 181 ARNTEFQTKDKALAEIVNHLGIIIVGKDGSVNLIGKVGKNEGYSVNGGSIISLLAGOKIT 240
 Qy 241 ISDIINPTIYSIAAPENEAIVNLGDIIFAKGGINVRAATIRNOQKLSADSVSKDKSGNIV 300
 Db 241 ISDIINPTIYSIAAPENEAIVNLGDIIFAKGGINVRAATIRNOQKLSADSVSKDKSGNIV 300
 Qy 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKGGEGTYLGGDERGE 360
 Db 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKGGEGTYLGGDERGE 360
 Qy 361 GKNGIQLAKKTSLEKSGTINVSQKGGFAIVWGDIALIDGNINAQSGDIATKGGFVET 420
 Db 361 GKNGIQLAKKTSLEKSGTINVSQKGGFAIVWGDIALIDGNINAQSGDIATKGGFVET 420
 Qy 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGGINDEFTGTG-EASDPKKNSEL 479
 Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGGINDEFTGTG-EASDPKKNSEL 479
 Qy 480 KTLTNTTISNLYKNWMTNTASRKLTVNSSINIGSNLSHLILHSKGGRGVQIDGDIIT 539
 Db 479 KTLTNTTISNLYKNWMTNTASRKLTVNSSINIGSNLSHLILHSKGGRGVQIDGDIIT 537
 Qy 540 ----SKGNLTITSGGWVVDVHKNTLD-QGFNLITA-ASVAFEGGNKAKDAANAKIVAQ 593
 Db 538 TGDDTFRGANLTITSGGWVVDVHKNTLSLGAQGNINITAKODIAFERKSGNOV-----ITGQ 590
 Qy 594 GTVTITGEGKDFRANNVSLNGTGGKLNLISSVNN---LTHNLSGTINTISGNTITQTRK 650
 Db 591 GTIT-SGNQKGRFRNPNVSLNGTSGLOPQTTKRTNKYAITNKFEGTLNISGVNISMVLPK 649

Qy 651 NTS-YWQTS-HSDSHMNVSAINLETGANFTF-IKYISSNSKGLTTOYRSSAGVNFVNGVNM 708
 Db 650 NESGYDKFKGRTYWNLTSLNVSESEGFNLTIDSRGSDSAGTLTQPYNLNGISF---NKDT 706
 Qy 709 SPNLKGAQVNFKLPNEMNMTSKLP-IRELANITATGGGVFFDIYANISG---RGAE 764
 Db 707 TPNVERNARVNEDIKAPIGINKYSSLVASPMGNISVSGGSDFTLLASSNVOTPGV 766
 Qy 765 LKMSSEINTSNGANFTLNSHVHVGDDAFKINKDLTINATNSFSLRKTDDFDVGYARNAIN 824
 Db 767 INSKYFNVSTGSLRFKTSKTKTGFSEKDLTLNATGNTITLLQVEGT--DGMITGKGI 824
 Qy 825 STYNISILGNGVNTLGGQNSSSSITNITIEKAAANYTLFANNAPNOQNIIRDRVILKGLSLV 884
 Db 825 AKKNITFEGGNTIFGSRKAVTEIEGNVTINNANVTILGSDFDNHQ--KPLTIKKDVAIN 882
 Qy 885 NGSLSLTGADIKGNLTISESATFKGKTRDTLNTITNGTNNGTAEINITQGVKLGNTV 944
 Db 883 SCNLTAGGNIVNLAGNLTVESNANFKATNFTFNVGGLFDNKGNSNISIAKGARFKDID 942
 Qy 945 NDGDLNITTHAKRNORSIIIGDILINKKGSNLITDSNDAEIOIGNISOKEGNLTISDDK 1004
 Db 943 NSKNLSITINSSSTYRTIISGNITNKNGLDNLITNEGSDTEMOIGGDVSKOKEGNLTISDDK 1002
 Qy 1005 INITKOITIKKIDGEDSSSDATSNANLTIKTELKELFEDLSISGFNKAETAKDGRDLT 1064
 Db 1003 INITKOITIKAGVDGSDSDATNNANLTIKTELKELTODLNISGFNKAETAKDGSDLT 1062
 Qy 1065 IGNSDNGSGAEAKVTTFNNVKDSKISADGHVNTLNSKVKTSSSNGGREGNSDNDTGLTI 1124
 Db 1063 IGNTNSAD-GTNAKVTNFQVQDKSISADGHVNTLNSKVKTSSSNGGREGNSDNDTGLTI 1121
 Qy 1125 TAKNVEVNDKITSKLVNITA-SEKVTYTAGSTINATNGKASITTKT----- 1170
 Db 1122 DAKNVTNNNITSHKAVSISATSGETTKTGTINATTGNVEITAGTSGILGGIESSGS 1181
 Qy 1171 -----GDISGTISGNT 1181
 Db 1182 VTLTATEGALAVSNISGNTVTVTANSALTLTLAGSTIKGTESVTTSSQSGIDGGTISGTT 1241
 Qy 1182 VSVASVTDLTTSKSGIEAKSGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1241
 Db 1242 VEKATESLTTSQSSNSKIKATTCGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1301
 Qy 1242 ATEGAATLTGNTLTEAGSSITSTPKGOVDLLAONGSTAGSTINANVTNTTGLTTLTVA 1301
 Db 1302 ATEGAATLTSSCKLTTEASSHTSAGQVNLQAQDSVAGSINANVTNTTGLTTLTVK 1361
 Qy 1302 GSDIKATSGTLVINAKDAKNGDASGDSTEVAVNAGSGSVTAATSSSVNTGDLNTVN 1361
 Db 1362 GSNINATSGTLVINAKDAKNGDASGDSTEVAVNAGSGSVTAATSSSVNTGDLNTVN 1421
 Qy 1362 GLNIIISKDGRNTVRLRKEIEVKYIQPVASVEVEVTEAKRVLEKVKDLSDERETLAKUG 1421
 Db 1422 GLNIIISKDGRNTVRLRKEIEVKYIQPVASVEVEVTEAKRVLEKVKDLSDERETLAKUG 1481
 Qy 1422 VSAVREVEPNNTTIVTQNEFTTRPSSQVITSEKACFSGSGNARVCTVNAVDGQ 1476
 Db 1482 VSAVREVEPNNTTIVTQNEFTTRPSSQVITSEKACFSGSGNARVCTVNAVDGQ 1536

RESULT 12

AAB01846

ID AAB01846 standard; Protein; 1536 AA.

XX AAB01846;

XX AAB01846;

DT 11-SEP-2000 (first entry)

XX Haemophilus influenzae strain 12 HmW1A protein, SEQ ID NO:67.

XX HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

XX non-typeable Haemophilus influenzae; NTH1; non-encapsulated;

KW

recombinant production; *Escherichia coli*; antibacterial; vaccine;
human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
detection; diagnosis.

Haemophilus influenzae strain 12.

WO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

07-OCT-1998; 98US-0167568.

08-DEC-1998; 98US-0206942.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-303789/26.

N-PSDB; AAA52195.

Nucleic acid molecule for producing recombinant high molecular weight
proteins of *Haemophilus* which are used as a vaccine to provide
protection against *Haemophilus* induced diseases in humans -

Example 16; Fig 28A-Q; 307pp; English.

The invention relates to the recombinant production of *Haemophilus*
influenzae high molecular weight (HMW) proteins in *Escherichia coli*. The
expression construct used to effect recombinant expression comprises a
promoter functional in *E. coli* (e.g., the T7 promoter) operably linked
to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene
clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
and the hmwB and hmwC genes encode accessory proteins which are
responsible for post-translational processing and secretion of the HMWA
proteins. The modified hmwABC operon used in the expression construct of
the invention contains an A gene modified such that it encodes only the
mature HMWA. The invention also discloses hmwa genes (AA52175-AA52198)
strains Joyce, K1, A21, LDCD2, PMH1, 15 and 12. The nucleic acids and
vectors are used for the production of recombinant H. *influenzae* HMW
and HMWA proteins (AAB01824-B01849) from the non-typeable H. *influenzae*
strains Joyce, K1, A21, LDCD2, PMH1, 15 and 12. The nucleic acids and
vectors are used for the production of recombinant H. *influenzae* HMW
proteins which can be used as vaccines to mediate a humoral or
cell-mediated immune response to provide protection against diseases in
humans caused by H. *influenzae* (e.g., otitis media, epiglottitis,
pneumonia and tracheobronchitis). The HMW proteins are also useful as
antigens in immunoassays for detecting antibodies against *Haemophilus*,
HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
HMW proteins can be used to isolate and clone hmw genes from other
non-typeable strains of *Haemophilus* via hybridisation reactions. The
present sequence represents an HMWA protein from a non-typeable strain of
H. *influenzae*.

SQ Sequence 1536 AA;

Query Match 66.7%; Score 4943; DB 21; Length 1536;

Best Local Similarity 67.1%; Pred. No. 5.4e-227;

Matches 1043; Conservative 147; Mismatches 267; Indels 98; Gaps 20;

QY 1 MNKTYRLKSKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLSSGVT 60

DB 1 MNKTYRLKSKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLSSGVT 60

QY 61 SIPOSVLASGLQMDVVHGTATWQVDGNKTIIRNSVDIINWKQFNIDQNMWQVLOENN 120

DB 61 SIPOSVLASGLQMDVVHGTATWQVDGNKTIIRNSVDIINWKQFNIDQNMWQVLOENN 120

QY 121 NSAVFNRTSNQISQLKGLDSNGOVFLNPNNGITIGKDALINTNGFTASTLDSISNENIK 180

DB 121 NSAVFNRTSNQISQLKGLDSNGOVFLNPNNGITIGKDALINTNGFTASTLDSISNENIK 180

QY 181 ARNFTTEQTKD KALAEIVNHLITVKGDSVNLIGKVKNEGVSIVNGGSISLLAGOKIT 240
DB 181 ARNFTTEQTKD KALAEIVNHLITVKGDSVNLIGKVKNEGVSIVNGGSISLLAGOKIT 240
QY 241 ISDIINPTTYSIAAPENEAVALNGDIFAKGGINVRAATIRNOGKLSADSVSKDGSNIY 300
DB 241 ISDIINPTTYSIAAPENEAVALNGDIFAKGGINVRAATIRNOGKLSADSVSKDGSNIY 300
QY 301 LSAKEGEAEIGGVISAQNOQAKGGLMTITGDKVTLTKTGAVIDLSGKEGETYLGGBERGE 360
DB 301 LSAKEGEAEIGGVISAQNOQAKGGLMTITGDKVTLTKTGAVIDLSGKEGETYLGGBERGE 360
QY 361 GKNGIOLAKKTSLKSGSTINVSCKEKGGAIVMGDIALIDGNINAQSGDIAKTGFVET 420
DB 361 GKNGIOLAKKTSLKSGSTINVSCKEKGGAIVMGDIALIDGNINAQSGDIAKTGFVET 420
QY 421 SCHDLFIKDNAIVDAKEWLLDFDNVSNINAEADPLFNNTGINDPEPTGTG-EASDPKKNSEL 479
DB 421 SCHDLFIKDNAIVDAKEWLLDFDNVSNINAEADPLFNNTGINDPEPTGTG-EASDPKKNSEL 479
QY 480 KTTLTNTTISNYLKNAWMTNITASRKLTVNSSINIGSNHSLHLHSGKORGQGVQIDGDIT 539
DB 479 KTTLTNTTISNYLKNAWMTNITASRKLTVNSSINIGSNHSLHLHSGKORGQGVQIDGDIT 539
QY 540 ----SKGNLTIYSGGWDVHKNTILD-QGFLNITA-ASVAFEGGNKKAARDAAAKIVAQ 593
DB 538 TGDDTRGANLTIYSGGWDVHKNTISLGAQGNINITAKQDIAFEKGSNQV-----ITGQ 590
QY 594 GTVTITGCKDPRANNVSLUNGTKGLNISSVNN---LTHNLSGTINISGNITINOTTRK 650
DB 591 GTIT-SGNQKGRFRNNVSLUNGTKGLQFTTKTKNYAITNKEGTLNIGKVNISMVLPK 649
QY 651 NTS-YWQTSHDSSHNVSAALNLETGANFTF-IKYISSNKGKLTQYRSSAGVNFVNGNM 708
DB 650 NESGYDKFKGRTYWNLTSLNVSESEFNLIDSRGSDSAGTLTQPVNLNGISF---NKDT 706
QY 709 SFNLKEGAKVNFKLKPNENMNTSKPLP-IRFLANTATGGGVSFFDIYANHSG---RGAE 764
DB 707 TFNVERNARVNFIDKAPGINKYSLNYSFNGNSVSGGSDVDTLLASSNVQTPGVV 766
QY 765 LKMSINISNGANFTLNSHVRGDDAFKINKDLTINATSNFSLRQTKDDFYDGYARNAIN 824
DB 767 INSKYFNVSTGSSURFKTSGSKTKGFSLEKDLTLNATGNTLLQVEGT--DGMITGKGLV 824
QY 825 STYNISILGQNVTLGGQSSSITGNITIEKAANYTLEANNAPNOQNRIDRVIKGLSLIV 884
DB 825 AKKNIIEGGINITFGSRKAVTEIEGNTINNANNVTLIGSDFDNHQ---KPLTKKDVIIIN 882
QY 885 NGSLTGCENADIKGNLTISESATFKGKTRDTLNTGTNNGTAEINITQGVVKGWVT 944
DB 883 SGNLTAGNIVNIAAGNLTVESNANFKAITNFTNVGGLDFDNKGNISIAKGGARFKDID 942
QY 945 NGDLNITTHAKRNQRTIIGDIIINKKGLSLNITDSNDAEIQIGNISKQECNLTISDDK 1004
DB 943 NSKNLSITTSNSTYRTIISGNITNKGDLNITNEGSDTEMQIGDVSQKEGNLTISDDK 1002
QY 1005 INITKQITIKGIDCEBSSSDATSNANLITKTELKLTEDLSISGFNKAETAKDGRDLT 1064
DB 1003 INITKQITIKAGVDGSDSDATNNANLITKTELKLTQDLNISGFNKAETAKDGSDLT 1062
QY 1065 IGNSDNCNGAEATVFNNVKDSKISADGHVTVLNSKVKTSSSNGGRESNDNDTGLTI 1124
DB 1063 IGNTNSAD-GTNNAKVTFFNQVKDKISADGHKVTLSHKVETSGSNNNTEDSDNNAGLTI 1121
QY 1125 TAKNVEVNDITSLKTVNITA-SEKVTITTAGSTINAYNGKASITTKT----- 1170
DB 1122 DAKNVTNNITSHKAVSISATSGEITTKGTTINATGTVNVEITATGSLIGLGISSSGS 1181
QY 1171 -----GDISGTISGNT 1181
DB 1182 VTLTATEGALAVSNISGNTVTVTANSALTLTLAGSTIKGTESVTTSSQSGDIGTISGNT 1241
QY 1182 VVSATVDLTTKSGSKTEAKSGEANVT SATCTIGCTISGNTVNTVNTANAGDLTVGNGAEIN 1241

```

Db 1242 VEVKATESLTQSNKIKATTGEANVTATGTTIGTISGNTVNTANAGDLTVNGABIN 1301
QY 1242 ATEGAATLATGNTLTTEAGSSITSTKGQVLLAQAQSIAGSINAANYTLNLTGTLTTVA 1301
Db 1302 ATEGAATLTSSGKLTTEASSHTSAKQVNLQAQSGVAGSINAANYTLNLTGTLTTVK 1361
QY 1302 GSDIKATSTLVNNAKDLNDASGDSDEVAVNASGSGVTAATSSSVNITGDLNTVN 1361
Db 1362 GSNINATSTLVNNAKDAELNHNHTVNTNANGSGSVIATSSSRVNTIGDLNTIN 1421
QY 1362 GLNIIISKGRNTVRLRGKEIEVKYIOPGVASVEEVEAKRVLEKVKDLSDEERE*TLAKLG 1421
Db 1422 GLNIIISKNGINTVLLKGVKIDVKYIOPGIVASVDEVEAKRVLEKVKDLSDEEREALAKLG 1481
QY 1422 VSAVRFEVGNNTITVNTONEFTTRPSOVIIISGKACFSSGNGARVCTNVADGQ 1476
Db 1482 VSAVRFEVGNNTITVNTONEFTTRPSOVIIISGKACFSSGNGARVCTNVADGQ 1536

```

RESULT 13

AAR41725
ID AAR41725 standard; Protein; 1536 AA.

XX AAR41725;

XX 26-APR-1994 (first entry)

XX High molecular weight protein 1 (HMW1).

XX HMW; high molecular weight protein; virus; vaccine; influenza;

XX epitope; immunity; haemophilus influenzae; gene cluster.

XX Haemophilus influenzae.

XX Key Location/Qualifiers

XX Misc-difference 668..677
XX /note= "Possibly incorrect sequence. Alternative
XX sequence for this region is LNVSESGFN.
XX (See comments)."

XX WO9319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993; 93WO-US02166.

XX 16-MAR-1992; 92GB-0005704.

XX (BARE/) BARENKAMP S J.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

XX N-PSDB; AQA49508.

XX High molecular weight surface proteins - of non-typeable
XX haemophilus which exhibit immunogenic properties

XX Claim 3; Figure 2/10; 100pp; English.

XX The isolation and purification of the high molecular weight protein
XX enables the identification of the major protective epitopes of the
XX protein by conventional epitope mapping. These epitopes can then be
XX synthesised using standard techniques and incorporated into fully
XX synthetic or recombinant vaccines. This sequence is claimed to be
XX the same as that given in AAR41723 (High molecular weight protein 1)
XX although it does differ slightly. (Repeated regions which are
XX possibly incorrect and occur in the corresponding nucleotide coding
XX sequence contribute to these differences).

XX Sequence 1536 AA;

SQ

```

Query Match 66.5%; Score 4924; DB 14; Length 1536;
Best Local Similarity 67.0%; Pred. No. 4.3e-226;
Matches 1042; Conservative 147; Mismatches 268; Indels 98; Gaps 20;

QY 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSALLSLGVT 60
Db 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSALLSLGVT 60
QY 61 SIPOSVLASGLQGMVHVHTATMQVDGNKTIIRNSVDALINWKOFNDQNMVQFLOENN 120
Db 61 SIPOSVLASGLQGMVHVHTATMQVDGNKTIIRNSVDALINWKOFNDQNMVQFLOENN 120
QY 121 NSAVFNRVTSNOISQLKGLDSNGOVFLNPGITIGKDAINTNGFTASTLDISNENIK 180
Db 121 NSAVFNRVTSNOISQLKGLDSNGOVFLNPGITIGKDAINTNGFTASTLDISNENIK 180
QY 181 ARNFTFEQTKDALKAEIVNHGLITVKGDSVNLIGGKVKNEGIVSVANGSISLLAGOKIT 240
Db 181 ARNFTFEQTKDALKAEIVNHGLITVKGDSVNLIGGKVKNEGIVSVANGSISLLAGOKIT 240
QY 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSKDSGNIV 300
Db 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSKDSGNIV 300
QY 301 LSAKEGEAEIGVISAQNOQAQKGLMITGDKVTLTKGAVTDLGSKGEGEYTLGGDERGE 360
Db 301 LSAKEGEAEIGVISAQNOQAQKGLMITGDKVTLTKGAVTDLGSKGEGEYTLGGDERGE 360
QY 361 GRNGIOLAKKTSLEKGSINVSKEKGGFAIWWGDIALIDGNIINAGSGSDIAKTGGFVET 420
Db 361 GRNGIOLAKKTSLEKGSINVSKEKGGFAIWWGDIALIDGNIINAGSGSDIAKTGGFVET 420
QY 421 SGHDLFIKDNATVDAKWLDPDNVNSINAEPLFNNTGINDDEFPTGTC-EASDPPKNSSEL 479
Db 421 SGHDLFIKDNATVDAKWLDPDNVNSINAEPLFNNTGINDDEFPTGTC-EASDPPKNSSEL 479
QY 480 KTTLTNTTISNLYKNAMTNITASKRLTVNNSINTGNSHLILHKSQGGVQDIDGIT 539
Db 480 KTTLTNTTISNLYKNAMTNITASKRLTVNNSINTGNSHLILHKSQGGVQDIDGIT 539
QY 539 KTTLTNTTLESILKKGTFVNITANORIYVNSINL-SNGSLTLWSEGRSGGVEINNDIT 537
Db 539 KTTLTNTTLESILKKGTFVNITANORIYVNSINL-SNGSLTLWSEGRSGGVEINNDIT 537
QY 540 ----SKGNNLIYSGGVVDVHKNIITLD-OGFLNITA-ASVAFEGGNNKARDAANAKIVAQ 593
Db 540 ----SKGNNLIYSGGVVDVHKNIITLD-OGFLNITA-ASVAFEGGNNKARDAANAKIVAQ 593
QY 593 TGDDTRGANLTIYSGGVVDVHKNIISLGAQGNINITAKQDIAFERGSNQV-----ITGO 590
Db 593 TGDDTRGANLTIYSGGVVDVHKNIISLGAQGNINITAKQDIAFERGSNQV-----ITGO 590
QY 594 GTVTTITGEGKDFRANNVSLNGTGKGLNIISVNN--LTHNLSGTINISGNITINQTRK 650
Db 594 GTVTTITGEGKDFRANNVSLNGTGKGLNIISVNN--LTHNLSGTINISGNITINQTRK 650
QY 591 GTIT-SGNQKGFERNVSLNGTSGLOFTTKRTNKYAITNKPEGLNISGKVNISMVLPK 649
Db 591 GTIT-SGNQKGFERNVSLNGTSGLOFTTKRTNKYAITNKPEGLNISGKVNISMVLPK 649
QY 651 NTS-YWOTSHDSHNV-SALNLETGANFTFIKYISSNKGTLTOYRSSAGVNFVNGVNM 708
Db 651 NTS-YWOTSHDSHNV-SALNLETGANFTFIKYISSNKGTLTOYRSSAGVNFVNGVNM 708
QY 650 NESGYDKFKGRTYWNLTSKVDMINSKDALTIDSRGSDSAGTLTPYNLNGISF--NKDT 706
Db 650 NESGYDKFKGRTYWNLTSKVDMINSKDALTIDSRGSDSAGTLTPYNLNGISF--NKDT 706
QY 709 SPNLKEGAKVNFELKPNENWNTSKPLP-IRFLANITATGGSVFFDIYANHS--RGAE 764
Db 709 SPNLKEGAKVNFELKPNENWNTSKPLP-IRFLANITATGGSVFFDIYANHS--RGAE 764
QY 707 TPNVERNARVNDIKAPIGINKYSSLVYASFNNGISVSGGSDFTLLASSNVQTPGV 766
Db 707 TPNVERNARVNDIKAPIGINKYSSLVYASFNNGISVSGGSDFTLLASSNVQTPGV 766
QY 765 LRKSEINISGANFTLNSHVRGDDAPKINKDLTINATNSFNLSRQKDFDGYARNAIN 824
Db 765 LRKSEINISGANFTLNSHVRGDDAPKINKDLTINATNSFNLSRQKDFDGYARNAIN 824
QY 767 INSKYFNVTGSSSLRFTSGSTKTFGSIKEDLTNATGGNTITLLQVEGT--DMGIGKGLV 824
Db 767 INSKYFNVTGSSSLRFTSGSTKTFGSIKEDLTNATGGNTITLLQVEGT--DMGIGKGLV 824
QY 825 STYNISILGNNVTLLGQNSSSSITGNITIEKAAVNTLEANNAPNOQNIIDRVIKGSLV 884
Db 825 STYNISILGNNVTLLGQNSSSSITGNITIEKAAVNTLEANNAPNOQNIIDRVIKGSLV 884
QY 885 NGSLISITGENADIKGNLTISESATFKGKTRDTLUNGTNNGTAEINTQGVVLGNVT 944
Db 885 NGSLISITGENADIKGNLTISESATFKGKTRDTLUNGTNNGTAEINTQGVVLGNVT 944
QY 883 SGNLTAGGNIVNIAGNLTVESNANFAITNFTENVGGLFDKNGKNSISAKGAREFKID 942
Db 883 SGNLTAGGNIVNIAGNLTVESNANFAITNFTENVGGLFDKNGKNSISAKGAREFKID 942
QY 945 NQDGLNITTHAKRNORSIIGGDIINKKGSUNITDSSNDAEIQIGGNISQKEGNLTSSDK 1004
Db 945 NQDGLNITTHAKRNORSIIGGDIINKKGSUNITDSSNDAEIQIGGNISQKEGNLTSSDK 1004
QY 943 NSKNLSITSSSTYRTIISGNITNKNGLNITNEGSDTEMQIGGVDVQKEGNLTSSDK 1002
Db 943 NSKNLSITSSSTYRTIISGNITNKNGLNITNEGSDTEMQIGGVDVQKEGNLTSSDK 1002

```

QY 1005 INITKQITIKKIDGESSDASNTIKTKELKLTEDLSISGFNKAEITAKGDRDIT 1064
 Db 1003 INITKQITIRAGVDGSSDSATNNANLITKTKELKLTQDLNLSGFNKAEITAKGDSLT 1062
 QY 1065 IGNSDNGSAGAEKVTFFNNVKDSKISADGHNTVLSKYTKSSNGGRESNDNDTGLTI 1124
 Db 1063 IGNTNSAD-GTNNAKVTFFNOVKDSKISADGHKVTLSKVKETSGSNNNTEDSDNNAGLTI 1121
 QY 1125 TAKNVEYNDKITSIKVTNITA-SEKVTYTAGSTINATNGKASITTKT----- 1170
 Db 1122 DAKNVTNNITSHKAVSISATSGEITTKTGTINATTGNVEITAGTSGILGIESSSGS 1181
 QY 1171 ----- 1171
 Db 1182 VTLTATEGALAVNISGNTVTVTANSALTLTLAGSTIKETESVTSSQSGDGGTISGCT 1241
 QY 1182 VVSATVADLTTRKSGKIEAKSGEANTVSATGTIGTISGNTVNVNANAGDLTVNGAEIN 1241
 Db 1242 VEKATESUTQTSNSKIKATTEGEANTVSATGTIGTISGNTVNVNANAGDLTVNGAEIN 1301
 QY 1242 ATEGAATLTATGNTLTTEAGSSTSTKGVLDLLAONGSTAGSINAAVTLNNTGTTLTVA 1301
 Db 1302 ATEGAATLTSSGKLTTEASSHITSAKQCVNLSAODGSVAGSINAANVTNLTCTLTIVK 1361
 QY 1302 GSDIKATSTLVINAKDAGKLDGSDSTEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1361
 Db 1362 GSNINATSGTLVINAKDAELNAGLNTVNVNATNANGSGSVIATTSRVNITGDLITIN 1421
 QY 1362 GLNIISKGRNVRGRGIEVKYTOPGVASVEEIVAEKRVLEKVDLSDEERETLAKLG 1421
 Db 1422 GLNIISKGINVLKGVKIDVYTOPGIVASVDEIVAEKRVLEKVDLSDEERETLAKLG 1481
 QY 1422 VSAREFVEPNNTITVNTONEFTTRPSSOVITISEGKACFSSGNGARVCTNVNADGQ 1476
 Db 1482 VSAREFIEPNNTITVDTONEATRLPSRIVISEGRACFSSDGTATVCVNIADNGR 1536

RESULT 14
 AAW30291
 ID AAW30291 standard; Protein; 1598 AA.
 XX
 AC AAW30291;
 XX
 DT 14-APR-1998 (first entry)
 XX
 DE Non-typeable Haemophilus high mol.wt. surface protein HMW3.
 KW Non-typeable Haemophilus; high molecular weight surface protein;
 KW HMW3; immunogen; vaccine; otitis media.
 XX
 OS Haemophilus influenzae strain 5.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 113 /note= "encoded by GTG"
 FT Misc-difference 864 /note= "encoded by TGT"
 XX
 PN WO9736914-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 01-APR-1997; 97WO-US04707.
 XX
 PR 01-APR-1996; 96US-0617697.
 XX
 PA (BARE/) BARENKAMP S J.
 XX
 PI Barenkamp SJ;
 XX
 PI 1997-503038/46.
 DR N-PSDB; AAT90992.
 XX

High molecular weight proteins of non-typeable Haemophilus influenzae - useful for vaccine production

Claim 1; Page 93-97; 183pp; English.

This protein comprises the high molecular weight surface protein HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that has the immunological ability to protect against disease caused by a non-typeable Haemophilus strain and is characterised by at least one surface-exposed B-cell epitope that is recognised by monoclonal antibody AD6. The HMW3 amino acid sequence was deduced from an isolated hmw3 gene (see AAT90992). HMW1 (see AAW30293), HMW2 (see AAW30294) and HMW4 (see AAW30292) have also been identified. A conjugate comprising HMW3 linked to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150 amino acids corresponding to at least protective epitope of HMW3 are also claimed. HMW proteins, conjugates and peptides can be used in vaccines, as immunogens for preparation of antibodies and as antigens for detection of these antibodies.

Sequence 1598 AA;

Query Match 64.1%; Score 4748.5; DB 18; Length 1598;
 Best Local Similarity 62.5%; Pred. No. 1e-217;
 Matches 1019; Conservative 143; Mismatches 281; Indels 187; Gaps 25;

QY 1 MNKIYRLKFKSRNLNVALVAVSELARGCDHSTKGESEKPMKVRHLAKPLSALLSLGVT 60
 Db 1 MNKIYRLKFKSRNLNVALVAVSELARGCDHSTKGESEKPMKVRHLAKPLSALLSLGMA 60
 QY 61 SIPOSVLASLQGMVHVHGTATMQVDGNKTIIIRNSVDIAIINWKQFNIDQNMVQFQENN 120
 Db 61 SIPOSVLASLQGMVHVHGTATMQVDGNKTIIIRNSVDIAIINWKQFNIDQNMVQFQLESS 120
 QY 121 NSAVENRVTNSQLSKGLDLSNGQVFLNPGITTKGDAIINTNGFTASTLDSINENIK 180
 Db 121 NSAVFNRVTSDQISQLKGLDLSNGQVFLNPGITTKGDAIINTNGFTASTLDSINENIK 180
 QY 181 ARNFTFEQTKDAEIVNHLITVVGKDSVNLIGKVKNEGVSIVNGSGISILAGOKIT 240
 Db 181 ARNFTLEQTKDAEIVNHLITVVGKDSVNLIGKVKNEGVSIVNGSGISILAGOKIT 240
 QY 241 ISDIINPTITYSTAAPENAEAVNLGDIKAGGNNINRAATIRNOGLSADSVSKDSGNIV 300
 Db 241 ISDIINPTITYSTAAPENAEAVNLGDIKAGGNNINRAATIRNOGLSADSVSKDSGNIV 300
 QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGETYLGDERGE 360
 Db 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGETYLGDERGE 360
 QY 361 GKNGIOLAKKTSLEKGSTINVSKEKGGFAIWMGDIALIDGNINAQGGDIAGTGGFVET 420
 Db 361 GKNGIOLAKKTTLEKGSTINVSKEKGGFAIWMGDIALIDGNINAQGGDIAGTGGFVET 419
 QY 421 SGHDLFIKDNAIVDAKEWLLDFDNVSINAEOLPNNNTGIDNDEFFPTGTGEADSP--KKNSE 478
 Db 420 SGHYSIDDNAIVTKEWLLDPENVITIEA--PSASRVELGADRNSHSAEVKVTLLKNNKT 477
 QY 479 LKTLTNTTISNYLKNWMTMTASRKLTVNASSINIGSNHLLHLSKGGGGVQIDGDI 538
 Db 478 SLTTLTNTTISNLLKSAHVNVITAPRKLTVNASSISIERGSHLLHSEGGGGVQIDKDI 537
 QY 539 TSKGNLTIYSGGVVDVHKNTLDOGFNITA--ASVAFE--GNNKARDANAANAKIYAQG 594
 Db 538 TSEGNLTIYSGGVVDVHKNTLDOGFNITA--ASVAFE--GNNKARDANAANAKIYAQG 591
 QY 595 TVTITGEGKDFRANVSLNGTGKGLNISS-----VNNLTNLSGTINISNITINQ 646
 Db 592 TIT-SGNSNGFRFNNVSLNSLGGKLSFTDSREDRGRRTKGNISKNKFDGTNLISGTVDISM 650
 QY 647 TTRKNTSWQTSKSHNVNVSALNLETGANFTF-IKYISSNSKGLTQYRSAGVNFNGVN 705
 Db 651 KAPKVSWFYRDKGRTYWNTTLLNVTSGSKFNLSIDTSGSGTGPSIRNAELNGITFN--- 707

Qy	726	ENMUTSKPLRPLAN-----ITATGGSYFFDIYAHSG---RGAELKMEINISNG	775
Db	465	-----ASIMPFKNANYALFNEIDISVSGGSVNFKLNASSNIQTPGVIKSNQFNVS	519
Qy	776	ANFTLNSHVGDARINKDLTINATNSNFSLRQTKDDFYDGYARNAINSTYINISILG	835
Db	520	STLNLKAGSETAFSIENDLNLNATGGNITIROVEGT--DSRYNKVGAAKNITPKG	577
Qy	836	VTLGGQNSSITGNITIEKAANVTLE-ANNAPOQ---NIRDVRVILKSLLLVNGSL	892
Db	578	ITFSQKATTEIKGNVTINKNTNATLARGANFAENKSPINI-----AGNVINGN	631
Qy	893	ENADIIGNLTISESATFGKTRDTLIINTGNITNGTAEINTQOVVKLGWNTDGLNIT	952
Db	632	SIINIAGNLTVSKGMLQAITWTFNVAAGSDNNAGNISIARGAKFKDINTTSLNIT	691
Qy	953	THAKRNORSIIIGDIIINKKGLNITSDSNDAIEIQTGGNISOKENGLTISDDKINIK	1012
Db	692	TNSDTTYRTIIRKGNISNKSGDLNIIDKKSDAEIOIGGNISOKENGLTISDDKINIK	751
Qy	1013	IKKGIDGEDSSDATSNANLTIKKELKLTDELLISGFENKAEITAKDGRDLTIGNS	1072
Db	752	IKAGVEGGSDSSEAEANLTIOTKELAGDLNISGFENKAEITAKNGSLDTIGNASG	811
Qy	1073	SGAEAKTVFNNVKDKISADCHNVTLNSKVKTSSNGGRESNDNDTGLTITAKNVE	1132
Db	812	--ADAKKVTFDKVKDKISTDGHVTLNSEVKT--SNGSSNAGNDNSTGLTISAKDV	867
Qy	1133	KDITSLKTVNI-----TASEKVT	1150
Db	868	NVTSHTKINTISAAAGNVTTKEGTTINATTGSVEVTAQNGTIKSNITQSNVTVT	927
Qy	1151	TTAGSTINATNGKASITTKTGDISG-----TISGNTVSVS-----ATVD--	1189
Db	928	TTENAVNATSGTVNISFTKDIKGGIESTSGNVNITASGNTLKVSNITGQDVTY	987
Qy	1190	-LTTKSGSIEAKSGEANTVSATG-----	1212
Db	988	ALTTTAGTISATTGNANITTKTGIDNGKVESGSVTLVATGATLAVGNISGNTV	1047
Qy	1213	-----TIG-----QTISGNTVNVATANAGDLTVCGNAEINATGCA	1246
Db	1048	DSGKLTSTVGSTINGTNSVTTSSOSGDIETISGNTVNVTAAGDITGNSAKVBAK	1107
Qy	1247	ATLTAATGNLTLTEAGSSITSTKGQVDLLANGSIAGSINAAVNTLMTTGLTFTV	1306
Db	1108	ATPLTAESKGLTQTGCSITTSNGQTTLTAKDDSIAGINAAVNTLMTTGLTFTG	1167
Qy	1307	ATSGTLVINAADAKLNGDAGDSTEVNAVNASGGSVTAATSSSYNITGDLNVTN	1366
Db	1168	ATSGTLTINAKDALDCAASGDRTVNATNASSGNNVTAKTSSSYNITGDLNVTN	1227
Qy	1367	SKDGRNTVLRKGEIVKYIOPGVASVEEVIEAKRVLEKVKDLSDEERETLAKL	1426
Db	1228	SENGRNTVLRKGEIDVKYIOPGVASVEEVIEAKRVLEKVKDLSDEERETLAKL	1287
Qy	1427	FVEPNNTITVNTONEFTTRPSSQVITSEKACFSSGNGARVCTNVADQ	1476
Db	1288	FVEPNNAITVNTONEFTTKPSQVITSEKACFSSGNGARVCTNVADQ	1337

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: March 24, 2003, 15:24:19 ; Search time 18 Seconds
(without alignments)
2414.314 Million cell updates/sec

Title: US-10-092-880-4

Perfect score: 7407

Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7407	100.0	1477	1	US-08-038-682-4
2	7407	100.0	1477	1	US-08-302-832-4
3	7407	100.0	1477	2	US-08-530-198-4
4	7407	100.0	1477	2	US-08-469-880-4
5	7407	100.0	1477	2	US-08-728-470-4
6	7407	100.0	1477	2	US-08-617-697-4
7	7407	100.0	1477	4	US-08-719-641-4
8	7261	98.0	1477	4	US-09-206-942-71
9	5282	71.3	1600	2	US-08-617-697-10
10	5164	69.7	1036	4	US-09-206-942-73
11	4967	67.1	1536	1	US-08-038-682-2
12	4967	67.1	1536	1	US-08-302-832-2
13	4967	67.1	1536	2	US-08-530-198-2
14	4967	67.1	1536	2	US-08-469-880-2
15	4967	67.1	1536	2	US-08-728-470-2
16	4967	67.1	1536	2	US-08-617-697-2
17	4967	67.1	1536	4	US-08-719-641-2
18	4956	66.9	1529	2	US-08-728-470-10
19	4956	66.9	1529	4	US-08-719-641-10
20	4943	66.7	1536	4	US-09-206-942-67
21	4751.5	64.1	1599	2	US-08-617-697-9
22	3525.5	47.6	1338	2	US-08-728-470-9
23	3525.5	47.6	1338	4	US-08-719-641-9
24	3396.5	45.9	983	4	US-09-206-942-51
25	3365.5	45.4	977	4	US-09-206-942-53
26	3260	44.0	998	4	US-09-206-942-59
27	3256.5	44.0	1079	4	US-09-206-942-47

28	3229	43.6	992	4	US-09-206-942-61
29	3225.5	43.5	1073	4	US-09-206-942-49
30	3184	43.0	975	4	US-09-206-942-30
31	3153	42.6	969	4	US-09-206-942-32
32	2970.5	40.1	1011	4	US-09-206-942-39
33	2939.5	39.7	1005	4	US-09-206-942-41
34	2864	38.7	1010	4	US-09-206-942-55
35	2833	38.2	1004	4	US-09-206-942-57
36	2763	37.3	1095	4	US-09-206-942-69
37	2741	37.0	1101	4	US-09-206-942-43
38	2710	36.6	1095	4	US-09-206-942-45
39	2678.5	36.2	1226	4	US-09-206-942-26
40	2647.5	35.7	1220	4	US-09-206-942-28
41	2609.5	35.2	1180	4	US-09-206-942-65
42	2609.5	35.2	1188	4	US-09-206-942-63
43	2584.5	34.9	1228	4	US-09-206-942-34
44	2553.5	34.5	1222	4	US-09-206-942-37
45	2066.5	27.9	915	4	US-09-206-942-35

ALIGNMENTS

RESULT 1
US-08-038-682-4
; Sequence 4, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-682-4

Query Match 100.0%; Score 7407; DB 1; Length 1477;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSEKPAKMKVRLHAKPLSAMLISLQVT 60
Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSEKPAKMKVRLHAKPLSAMLISLQVT 60

QY 61 SIPOVSLASGLQMDVYHGTATQVQDNKTIIRNSVDALINNKQFNIDONEMVQFLOENN 120
Db 61 SIPOVSLASGLQMDVYHGTATQVQDNKTIIRNSVDALINNKQFNIDONEMVQFLOENN 120
QY 121 NSAVFNRTVSNQISQIKGLDSNGQVFLNPNIGITGKDAINTNGFRTASTLDISNENIK 180
Db 121 NSAVFNRTVSNQISQIKGLDSNGQVFLNPNIGITGKDAINTNGFRTASTLDISNENIK 180
QY 181 ARNFTFEOTKDKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSISLLAGQKIT 240
Db 181 ARNFTFEOTKDKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSISLLAGQKIT 240
QY 241 ISDIINPTITYIAAPENAEVNLGDFIAKGGNINRAATIRNQGKLSADSVSKDKSGNIV 300
Db 241 ISDIINPTITYIAAPENAEVNLGDFIAKGGNINRAATIRNQGKLSADSVSKDKSGNIV 300
QY 301 LSAKEGEAEIGVISAQNOQAQKGLMITGDKVTLTKTGAVIDLSKEGEGEYVLGGDERGE 360
Db 301 LSAKEGEAEIGVISAQNOQAQKGLMITGDKVTLTKTGAVIDLSKEGEGEYVLGGDERGE 360
QY 361 KNGIQLAKRTSLEKSTINVSKEKGGFAIVWGDIALIDGNIINAQSGDIAKTGGFVET 420
Db 361 KNGIQLAKRTSLEKSTINVSKEKGGFAIVWGDIALIDGNIINAQSGDIAKTGGFVET 420
QY 421 SGHDLFIKDAIVDAKEWLLDFDNVSIINAEDPLFNNTGIDEFPTGTGEASDPKKNSELK 480
Db 421 SGHDLFIKDAIVDAKEWLLDFDNVSIINAEDPLFNNTGIDEFPTGTGEASDPKKNSELK 480
QY 481 TTLNTTISNLYKNAWTNITASRLTWNSSINIGSNHLLIHSKGQGGVQIDGDTIS 540
Db 481 TTLNTTISNLYKNAWTNITASRLTWNSSINIGSNHLLIHSKGQGGVQIDGDTIS 540
QY 541 KGNLTIVSGWVDVHKNTLDQGFNLITAAVAFEGGNKARDAANAIAVAQGTVTITG 600
Db 541 KGNLTIVSGWVDVHKNTLDQGFNLITAAVAFEGGNKARDAANAIAVAQGTVTITG 600
QY 601 EKGDFRANVSLNGTGKLNLISSVNNLTHNLSGTINISGNITINOTTRKNTSYWQTSND 660
Db 601 EKGDFRANVSLNGTGKLNLISSVNNLTHNLSGTINISGNITINOTTRKNTSYWQTSND 660
QY 661 SHNVNLSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGNMSFNLKEGAKVNF 720
Db 661 SHNVNLSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGNMSFNLKEGAKVNF 720
QY 721 KLKPNENMTSKPLPIRPLANITATGGSVFFDIYANHSRGAEIKWSEINISNCAFTL 780
Db 721 KLKPNENMTSKPLPIRPLANITATGGSVFFDIYANHSRGAEIKWSEINISNCAFTL 780
QY 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNISILGNGVTLGG 840
Db 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNISILGNGVTLGG 840
QY 841 QNSSSTIGNTTIEKAANVTLEANNAPNQNRDVRKLGSLVNGSLTGENADIKGN 900
Db 841 QNSSSTIGNTTIEKAANVTLEANNAPNQNRDVRKLGSLVNGSLTGENADIKGN 900
QY 901 LTISESATFKGKTBDTLNITGNFTNNGTAENITQGVVVKLGNTVNDGDLNITTHAKRNQR 960
Db 901 LTISESATFKGKTBDTLNITGNFTNNGTAENITQGVVVKLGNTVNDGDLNITTHAKRNQR 960
QY 961 SIIGGDIINKKSLNITSDSNDAEIQIGGNTSQEGLNITSSDKINITKQITIKKGIDGE 1020
Db 961 SIIGGDIINKKSLNITSDSNDAEIQIGGNTSQEGLNITSSDKINITKQITIKKGIDGE 1020
QY 1021 DSSSDATSNANLTITKTELKLTEDLSISGFNKAIPITAKDGRDLTIGNSDNCSGAEAKTV 1080
Db 1021 DSSSDATSNANLTITKTELKLTEDLSISGFNKAIPITAKDGRDLTIGNSDNCSGAEAKTV 1080
QY 1081 TFNNVKDSKISADGHNVTLSKVKYKTSNNGSGRESNSDNDTGLTITAKNVEVNDKITSLEKT 1140
Db 1081 TFNNVKDSKISADGHNVTLSKVKYKTSNNGSGRESNSDNDTGLTITAKNVEVNDKITSLEKT 1140

QY 1141 VNITASEKYTTAGSTINATNGKASITTKTGDISTGTSIGNTVSVSATVLTITKSGSKIEA 1200
Db 1141 VNITASEKYTTAGSTINATNGKASITTKTGDISTGTSIGNTVSVSATVLTITKSGSKIEA 1200
QY 1201 KSGEANVTSATGTIGGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260
Db 1201 KSGEANVTSATGTIGGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260
QY 1261 GSSITSTKGQVDLLAQNGSIAGSINAAVNTLTGTLTTVAGSDIKATSGTLVINAKDAK 1320
Db 1261 GSSITSTKGQVDLLAQNGSIAGSINAAVNTLTGTLTTVAGSDIKATSGTLVINAKDAK 1320
QY 1321 LMGDASGDSSTEVNAYNAGSGSVTAATSSSVNTGDLNTVNGLNIIISKDGRNVTLRGKE 1380
Db 1321 LMGDASGDSSTEVNAYNAGSGSVTAATSSSVNTGDLNTVNGLNIIISKDGRNVTLRGKE 1380
QY 1381 IEVKIOPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFVFPNNTITVNTQN 1440
Db 1381 IEVKIOPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFVFPNNTITVNTQN 1440
QY 1441 EFTTRPSQVILISEGKACFSSGNGARVCTNVADDGQP 1477
Db 1441 EFTTRPSQVILISEGKACFSSGNGARVCTNVADDGQP 1477

RESULT 2

US-08-302-832-4
; Sequence 4, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-302-832-4

Query Match

100.0%; Score 7407; DB 1; Length 1477;

Tue Mar 25 08:37:45 2003

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MNKIYRLKSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLALPLSALLSLGVT	60
Db	1	MNKIYRLKSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLALPLSALLSLGVT	60
Qy	61	SIPOQVSLAGLQCMQDVHGATMVDGNKTIIRNSVDALINWKQFNIDONEMVQFLOENN	120
Db	61	SIPOQVSLAGLQCMQDVHGATMVDGNKTIIRNSVDALINWKQFNIDONEMVQFLOENN	120
Qy	121	NSAVENRVTNSOISQLGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDISNENIK	180
Db	121	NSAVENRVTNSOISQLGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDISNENIK	180
Qy	181	ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSSISLLAGQKIT	240
Db	181	ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSSISLLAGQKIT	240
Qy	241	ISDIINPTITYSTAAPENAEVNLGDIKAGGNINVRATIRNOGKLSADSVSKDKSGNIV	300
Db	241	ISDIINPTITYSTAAPENAEVNLGDIKAGGNINVRATIRNOGKLSADSVSKDKSGNIV	300
Qy	301	LSAKEGEAEITGGVISAQNOQAAGKGLMITGDKVTLKTGAVIDLSKEGETYVLGGDERGE	360
Db	301	LSAKEGEAEITGGVISAQNOQAAGKGLMITGDKVTLKTGAVIDLSKEGETYVLGGDERGE	360
Qy	361	GKNGIOLAKKTSLEKSTINVSKEKGGFAIVWGDIALIDGNINAQSGDIAKTGGFVET	420
Db	361	GKNGIOLAKKTSLEKSTINVSKEKGGFAIVWGDIALIDGNINAQSGDIAKTGGFVET	420
Qy	421	SGHDLFTKONAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDFPTGTGEASDPKKNSELK	480
Db	421	SGHDLFTKONAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDFPTGTGEASDPKKNSELK	480
Qy	481	TLTLNTTISNLYKNAMTNITASKRLTVNSSINIGSNHLLHLSKQORGQGVQIDGDITS	540
Db	481	TLTLNTTISNLYKNAMTNITASKRLTVNSSINIGSNHLLHLSKQORGQGVQIDGDITS	540
Qy	541	KGNLTIYSGGWVDVHKNTLDQGLNITAAASVAFEGGNKARDAANAKIVAQGTVTITG	600
Db	541	KGNLTIYSGGWVDVHKNTLDQGLNITAAASVAFEGGNKARDAANAKIVAQGTVTITG	600
Qy	601	EGKDFRANVSLNGTKGLNISSVNNLTHNLSTINISGNIITINOTTRKNTSYWQTSKD	660
Db	601	EGKDFRANVSLNGTKGLNISSVNNLTHNLSTINISGNIITINOTTRKNTSYWQTSKD	660
Qy	661	SHNVSALENGANFTFIKYSNSKGLTQYRSSAGVNFVNGMSEFNLEKAGAVNF	720
Db	661	SHNVSALENGANFTFIKYSNSKGLTQYRSSAGVNFVNGMSEFNLEKAGAVNF	720
Qy	721	KLKPNENMTSKPLPIRFLANITATGGGVFFDIYANHSGRGAEKMKSEINISNGANFTL	780
Db	721	KLKPNENMTSKPLPIRFLANITATGGGVFFDIYANHSGRGAEKMKSEINISNGANFTL	780
Qy	781	NSHVRGDDAFKINKDLTINATNSFSLQTKDDFYDGYARNAINSTYINISILGNNVTLG	840
Db	781	NSHVRGDDAFKINKDLTINATNSFSLQTKDDFYDGYARNAINSTYINISILGNNVTLG	840
Qy	841	QNSSSSTGNTTIEKAANVTLEANNAPNOQNIQVVKLGLSVNGSLTGENADITGN	900
Db	841	QNSSSSTGNTTIEKAANVTLEANNAPNOQNIQVVKLGLSVNGSLTGENADITGN	900
Qy	901	LTISEATFKGKTROTTLNITGNFTNNGTAEINTQGVVVKLGNVNDGDLNITTHAKRNOR	960
Db	901	LTISEATFKGKTROTTLNITGNFTNNGTAEINTQGVVVKLGNVNDGDLNITTHAKRNOR	960
Qy	961	SIIGDDIINKGSLNITDSNNDAEIQGGNISQKEGNTITSSDKINIKTKKGGIDGE	1020
Db	961	SIIGDDIINKGSLNITDSNNDAEIQGGNISQKEGNTITSSDKINIKTKKGGIDGE	1020
Qy	1021	DSSSDATSNANLTITKTELKLTEDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080

Db	1021	DSSSDATSNANLTITKTELKLTEDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080
Qy	1081	TFNNVKDSKISADCHNVTLSKVKTSNGSGRESNDNDTGLTITAKNNVFNKDIITSLKT	1140
Db	1081	TFNNVKDSKISADCHNVTLSKVKTSNGSGRESNDNDTGLTITAKNNVFNKDIITSLKT	1140
Qy	1141	VNITASEKVTITAGSTINATNGKASITTKTGDISTGTSIGNTVSVSATVDLTTKSGSKIEA	1200
Db	1141	VNITASEKVTITAGSTINATNGKASITTKTGDISTGTSIGNTVSVSATVDLTTKSGSKIEA	1200
Qy	1201	KSGEANTYSATGTIGTISGNTVNTANAGDLTVNGAGAEINATEGAATLTATGNLTTEA	1260
Db	1201	KSGEANTYSATGTIGTISGNTVNTANAGDLTVNGAGAEINATEGAATLTATGNLTTEA	1260
Qy	1261	GSSITSTKGQVDDLLAONGSIAGSINAANVTLTNTTGTTLTVAGSDIKATSGTLVINAKDAK	1320
Db	1261	GSSITSTKGQVDDLLAONGSIAGSINAANVTLTNTTGTTLTVAGSDIKATSGTLVINAKDAK	1320
Qy	1321	LNGDASGDSSTEVNNAVNASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTVRLRGKE	1380
Db	1321	LNGDASGDSSTEVNNAVNASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTVRLRGKE	1380
Qy	1381	IEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVNTQN	1440
Db	1381	IEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVNTQN	1440
Qy	1441	EFTTRPSSOVIISGKACFFSSGNGARVCTNVADDGQP	1477
Db	1441	EFTTRPSSOVIISGKACFFSSGNGARVCTNVADDGQP	1477

RESULT 3
US-08-530-198-4
; Sequence 4, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-4

Query Match 100.0%; Score 7407; DB 2; Length 1477;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKIYRLKSKRLNALVAVSELARGDHSSTEGSEKPAKMKVPHLALPLSAMLISLGVY 60
DB 1 MNKIYRLKSKRLNALVAVSELARGDHSSTEGSEKPAKMKVPHLALPLSAMLISLGVY 60

QY 61 SIPOSVLASGLOGMDVHVHTATMOVDGNTIIRNSVDALINMKQFNIDONEMVQFLQENN 120
DB 61 SIPOSVLASGLOGMDVHVHTATMOVDGNTIIRNSVDALINMKQFNIDONEMVQFLQENN 120

QY 121 NSAVENRVTSMQISQKGLDLSNGQVFLNPNNGITIGKDAIINTNGFTASTLDISNENIK 180
DB 121 NSAVENRVTSMQISQKGLDLSNGQVFLNPNNGITIGKDAIINTNGFTASTLDISNENIK 180

QY 181 ARNFTPEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSLSLLAGQKIT 240
DB 181 ARNFTPEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSLSLLAGQKIT 240

QY 241 ISDIINPTIYSAAPENAVNLGDIIPAKGGINVRAATIRNOGKLSADSVSKDGSNIV 300
DB 241 ISDIINPTIYSAAPENAVNLGDIIPAKGGINVRAATIRNOGKLSADSVSKDGSNIV 300

QY 301 LSAKEGEAEIGGVISAQNOQAKGGLMITGDKVTLTKGAVIDLSCGEGEYTLGGDERGE 360
DB 301 LSAKEGEAEIGGVISAQNOQAKGGLMITGDKVTLTKGAVIDLSCGEGEYTLGGDERGE 360

QY 361 KNGIQLAKKTSLEKGSTINVSKEKGGFAIWMGDIALIDGNIINAQSGDIAGTGGFVET 420
DB 361 KNGIQLAKKTSLEKGSTINVSKEKGGFAIWMGDIALIDGNIINAQSGDIAGTGGFVET 420

QY 421 SGHDLPIKDNAIVDAKEWLLDPDVSINAEPLFNNTGINDFPTGTGGEADPKKNSL 480
DB 421 SGHDLPIKDNAIVDAKEWLLDPDVSINAEPLFNNTGINDFPTGTGGEADPKKNSL 480

QY 481 TTLTNTTISNVLKNAWTNITASRKLTVNSNINSGNSHLILHSGRGQGGVQIDGITS 540
DB 481 TTLTNTTISNVLKNAWTNITASRKLTVNSNINSGNSHLILHSGRGQGGVQIDGITS 540

QY 541 KGNLTIIYSGGWVVKHNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQGTVTITG 600
DB 541 KGNLTIIYSGGWVVKHNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQGTVTITG 600

QY 601 ECKDFRANVSLNGTGKGLNIISSVNNLTHNLSTINISGNITINOTTRKNTSYWQTS 660
DB 601 ECKDFRANVSLNGTGKGLNIISSVNNLTHNLSTINISGNITINOTTRKNTSYWQTS 660

QY 661 SHWVNSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGMSFNLKEGAKVNF 720
DB 661 SHWVNSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGMSFNLKEGAKVNF 720

QY 721 KLPENNNNTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMSINISNGANFTL 780
DB 721 KLPENNNNTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMSINISNGANFTL 780

QY 781 NSHVGDGDAFKINKDLINATNSNFSURQTKDFYDGYARNAINSTNYNISILGCVTLGG 840
DB 781 NSHVGDGDAFKINKDLINATNSNFSURQTKDFYDGYARNAINSTNYNISILGCVTLGG 840

QY 841 QNSSSTTGNITTEKAANVTLEANNAPNOQNI RDRVILKGLSLVNGSLSLTGENADIKN 900
DB 841 QNSSSTTGNITTEKAANVTLEANNAPNOQNI RDRVILKGLSLVNGSLSLTGENADIKN 900

QY 901 LTISESATFKGTRDTLNTNGTNGTAETINTQGVKLGWNTDGDNLNITTHAKRNOR 960
DB 901 LTISESATFKGTRDTLNTNGTNGTAETINTQGVKLGWNTDGDNLNITTHAKRNOR 960

QY 961 SIIGGDIINKKSLNITDSNNDAAEIQIGGNISOKEGNLTISDDKINITKQITIKKGIDGE 1020
DB 961 SIIGGDIINKKSLNITDSNNDAAEIQIGGNISOKEGNLTISDDKINITKQITIKKGIDGE 1020

QY 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV 1080
DB 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV 1080

QY 1081 TFNNVKDSKISADGHNVTLSKVKTSNNGGREGSNDNDTGLTITAKNVEVKNKDITSLKT 1140
DB 1081 TFNNVKDSKISADGHNVTLSKVKTSNNGGREGSNDNDTGLTITAKNVEVKNKDITSLKT 1140

QY 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTVSNTVSVSATVDTLTKSGSKIEA 1200
DB 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTVSNTVSVSATVDTLTKSGSKIEA 1200

QY 1201 KSGEANTVSATGTTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLPT 1260
DB 1201 KSGEANTVSATGTTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLPT 1260

QY 1261 GSSITSTKGQVDDLLAQNAGSINAGSINAAVNTLTGTTTAVAGSDIKATSGTLVINAKDAK 1320
DB 1261 GSSITSTKGQVDDLLAQNAGSINAGSINAAVNTLTGTTTAVAGSDIKATSGTLVINAKDAK 1320

QY 1321 LNGDASGDSDEVNAVNASGSGSVTAATSSSVNTGDLTNGVNLIIISKDGRTVRLRGKE 1380
DB 1321 LNGDASGDSDEVNAVNASGSGSVTAATSSSVNTGDLTNGVNLIIISKDGRTVRLRGKE 1380

QY 1381 IEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVNTON 1440
DB 1381 IEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVNTON 1440

QY 1441 EFTTRPSQVILIEGKACFSNGGARVCTNVADGGQP 1477
DB 1441 EFTTRPSQVILIEGKACFSNGGARVCTNVADGGQP 1477

RESULT 4
US-08-469-880-4
; Sequence 4, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810									
TELEFAX: (703) 415-0813									
INFORMATION FOR SEQ ID NO: 4:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 1477 amino acids									
TYPE: amino acid									
STRANDEDNESS: single									
TOPOLOGY: linear									
US-08-469-880-4									
Query Match									
Best Local Similarity 100.0%; Score 7407; DB 2; Length 1477;									
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTBKSGSEKPAKMKVRLHALKPLSAMLLSLGV	60						
DB	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTBKSGSEKPAKMKVRLHALKPLSAMLLSLGV	60						
QY	61	SIQSVLASGLQGMVDVHGTATMQVDGNKTIIRNSVDALINMKQFNIDQNEWQFLOENN	120						
DB	61	SIQSVLASGLQGMVDVHGTATMQVDGNKTIIRNSVDALINMKQFNIDQNEWQFLOENN	120						
QY	121	NSAVFNRTSNOISQKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDSNENIK	180						
DB	121	NSAVFNRTSNOISQKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDSNENIK	180						
QY	181	ARNFTFQPKDALKALAEIVNHGLITVKGDSVNLIGKVKNEGIVSVNGGSI SLLAGQKIT	240						
DB	181	ARNFTFQPKDALKALAEIVNHGLITVKGDSVNLIGKVKNEGIVSVNGGSI SLLAGQKIT	240						
QY	241	ISDINPTTYSIAAPENAEVNLGDFAKGGNINVRATIRNOGLSADSVSKDKSGNIV	300						
DB	241	ISDINPTTYSIAAPENAEVNLGDFAKGGNINVRATIRNOGLSADSVSKDKSGNIV	300						
QY	301	LSAKEGEAIGVISAOQOAGKKLMTGDKVTLTKTGAVIDLSKEGETVVLGGERGE	360						
DB	301	LSAKEGEAIGVISAOQOAGKKLMTGDKVTLTKTGAVIDLSKEGETVVLGGERGE	360						
QY	361	GKNGIOLAKTSLEKSTINVSKEKGGFAIVMGDIALDGNINAGSGDIAKTGGFVET	420						
DB	361	GKNGIOLAKTSLEKSTINVSKEKGGFAIVMGDIALDGNINAGSGDIAKTGGFVET	420						
QY	421	SGHDLFIKONAIVDKAEWLLDFDNVSINAEEDPLFNNTGINDPEPTGTEASDPKKNSELK	480						
DB	421	SGHDLFIKONAIVDKAEWLLDFDNVSINAEEDPLFNNTGINDPEPTGTEASDPKKNSELK	480						
QY	481	TTLTNTTISNLYKNAMTNITASRKLTVNSSINIGSNHLLILHSGQORGGVQIDGDITS	540						
DB	481	TTLTNTTISNLYKNAMTNITASRKLTVNSSINIGSNHLLILHSGQORGGVQIDGDITS	540						
QY	541	KGNLTITSGGWVDVHKNTLDQGLNITAAASVAFEGGNNKARDAANAKIVAQGVTTITG	600						
DB	541	KGNLTITSGGWVDVHKNTLDQGLNITAAASVAFEGGNNKARDAANAKIVAQGVTTITG	600						
QY	601	EGKDFRANNVSLNGTCKGLNIISSVNNLTHNLSGTINISGNITINOTTRKNTSYWQTS	660						
DB	601	EGKDFRANNVSLNGTCKGLNIISSVNNLTHNLSGTINISGNITINOTTRKNTSYWQTS	660						
QY	661	SHNVNLSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGMNSFNLKEGAKVNF	720						
DB	661	SHNVNLSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGMNSFNLKEGAKVNF	720						
QY	721	KLKPENNMTSKPLPIREFLANITATGGSVFFDIYAHNSGRGAELKMSIINISNGANFTL	780						
DB	721	KLKPENNMTSKPLPIREFLANITATGGSVFFDIYAHNSGRGAELKMSIINISNGANFTL	780						
QY	781	NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYISILGCVNLTGG	840						
DB	781	NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYISILGCVNLTGG	840						
QY	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIQDRVTKLGSLLVNGSLTGTENADIKGN	900						
DB	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIQDRVTKLGSLLVNGSLTGTENADIKGN	900						

QY	901	LTISSEATFKGKTRDRLNITNGTAEINITQGVKLGNTDGLNITTHAKRNOR	960						
DB	901	LTISSEATFKGKTRDRLNITNGTAEINITQGVKLGNTDGLNITTHAKRNOR	960						
QY	961	SIIGGDIINKKGSNLITDSNNDAEIOIGGNI SOKEGNLTISSDKINIKITQITIKKGIDG	1020						
DB	961	SIIGGDIINKKGSNLITDSNNDAEIOIGGNI SOKEGNLTISSDKINIKITQITIKKGIDG	1020						
QY	1021	DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGNSGAETV	1080						
DB	1021	DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGNSGAETV	1080						
QY	1081	TFNNVKDSKISADGHNVTLNSKVKTSSNGGRESNDNDTGLTITAKNVVKNKDITSLKT	1140						
DB	1081	TFNNVKDSKISADGHNVTLNSKVKTSSNGGRESNDNDTGLTITAKNVVKNKDITSLKT	1140						
QY	1141	VNITASEKVTITAGSTINATNGKASITTKTGDLSGTISGNTVSVSATVDLTTKSGSKIEA	1200						
DB	1141	VNITASEKVTITAGSTINATNGKASITTKTGDLSGTISGNTVSVSATVDLTTKSGSKIEA	1200						
QY	1201	KSGEANTVSATGITIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA	1260						
DB	1201	KSGEANTVSATGITIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA	1260						
QY	1261	GSSITSTKGQVDDLLAQNAGSIAGSINAANVTNTTGTITTVAGSDIKATSGTLVINAKDAK	1320						
DB	1261	GSSITSTKGQVDDLLAQNAGSIAGSINAANVTNTTGTITTVAGSDIKATSGTLVINAKDAK	1320						
QY	1321	LNGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLTNTVNLNIIISKDGRTVRLRGKE	1380						
DB	1321	LNGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLTNTVNLNIIISKDGRTVRLRGKE	1380						
QY	1381	IEVKYIOPGVASVEVIEAKRVLEKVKDLSDEERETLAKLGVSARFVPEPNTTIVNTQN	1440						
DB	1381	IEVKYIOPGVASVEVIEAKRVLEKVKDLSDEERETLAKLGVSARFVPEPNTTIVNTQN	1440						
QY	1441	EFTTRPSQVITISSEKACFFSGNGARVCTNNVADGQP	1477						
DB	1441	EFTTRPSQVITISSEKACFFSGNGARVCTNNVADGQP	1477						

RESULT 5
US-08-728-470-4
; Sequence 4, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barankamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728.470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-4

Query Match 100.0%; Score 7407; DB 2; Length 1477;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKYLKPSKRLNALVAVSELARGCDHSTKSGSEKPAKRVHRLALKPLSAMLISLQVT 60
DB 1 MKKYLKPSKRLNALVAVSELARGCDHSTKSGSEKPAKRVHRLALKPLSAMLISLQVT 60
QY 61 SIQSVLASGLQMDVVGHTATQVGNKTIIRNSVDAILNKKQFNIDQNMWQFLOENN 120
DB 61 SIQSVLASGLQMDVVGHTATQVGNKTIIRNSVDAILNKKQFNIDQNMWQFLOENN 120
QY 121 NSAVFNRTVNSQISOLKGLDSNGQVFLNPNGITIGKDALINNGFTASTFLDLSNENIK 180
DB 121 NSAVFNRTVNSQISOLKGLDSNGQVFLNPNGITIGKDALINNGFTASTFLDLSNENIK 180
QY 181 ARNFTFQTKOKALAEIVNHGLIITVGKDSYNLIGKVKKEGVISVNGGSIISLLAGOKIT 240
DB 181 ARNFTFQTKOKALAEIVNHGLIITVGKDSYNLIGKVKKEGVISVNGGSIISLLAGOKIT 240
QY 241 ISDIINPTIYISIAAPENAEVNLGDIIFAKGGINVRAATIRNQGKLSADSVSKDKSGNIV 300
DB 241 ISDIINPTIYISIAAPENAEVNLGDIIFAKGGINVRAATIRNQGKLSADSVSKDKSGNIV 300
QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKGGETILGGDERGE 360
DB 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKGGETILGGDERGE 360
QY 361 KNGIQLAKKTSLEKGTINVSQKKEGKGFALVWGDIALIDGNINAQSGGDIAGTGGFVET 420
DB 361 KNGIQLAKKTSLEKGTINVSQKKEGKGFALVWGDIALIDGNINAQSGGDIAGTGGFVET 420
QY 421 SGHDLFIKDAIVDAKEWLLDFDNVSIADPFLNNTGINDPEPTGTGEASDPKKNSELK 480
DB 421 SGHDLFIKDAIVDAKEWLLDFDNVSIADPFLNNTGINDPEPTGTGEASDPKKNSELK 480
QY 481 TTLNTTISYLNKAWTMTNITASKRLTVNSNIGNSHLILHSKGQGGVQIDGDIITS 540
DB 481 TTLNTTISYLNKAWTMTNITASKRLTVNSNIGNSHLILHSKGQGGVQIDGDIITS 540
QY 541 KGNLTIYSGGWVDVHKNTLDOGFNLITAAASVAFEGGNKARDAANAKIVAQGTITING 600
DB 541 KGNLTIYSGGWVDVHKNTLDOGFNLITAAASVAFEGGNKARDAANAKIVAQGTITING 600
QY 601 EGKDFRANVNSLNGTKGLNITSSVNLTHNLSTGINSNITINQTRKNTSYWQTSHD 660
DB 601 EGKDFRANVNSLNGTKGLNITSSVNLTHNLSTGINSNITINQTRKNTSYWQTSHD 660
QY 661 SHNVNLSALNLETGANFTFKYIISNSKGLTQYRSSAGVFNFGVNGNSFNKKEGAKVNF 720
DB 661 SHNVNLSALNLETGANFTFKYIISNSKGLTQYRSSAGVFNFGVNGNSFNKKEGAKVNF 720
QY 721 KLPKNENMNTSKPLPIRFUANITATGGGVFFDIYANHSGRGAELKMSSEINISNCANFTL 780
DB 721 KLPKNENMNTSKPLPIRFUANITATGGGVFFDIYANHSGRGAELKMSSEINISNCANFTL 780

DB 721 KLPKNENMNTSKPLPIRFUANITATGGGVFFDIYANHSGRGAELKMSSEINISNCANFTL 780
QY 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNGVTLGG 840
DB 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNGVTLGG 840
QY 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIROVRKLGSLVNGSLTGENADIKGN 900
DB 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIROVRKLGSLVNGSLTGENADIKGN 900
QY 901 LTISESATFKGKTRDTLNTGNFTNNGTAEINITQGVVVKLCNVNTDGLNITTHAKRNOR 960
DB 901 LTISESATFKGKTRDTLNTGNFTNNGTAEINITQGVVVKLCNVNTDGLNITTHAKRNOR 960
QY 961 SIIGGDIINKKSLNITDSNNDAIEIQIGGNIQSGEGLTISSSDKINITKQITTKKGIDGE 1020
DB 961 SIIGGDIINKKSLNITDSNNDAIEIQIGGNIQSGEGLTISSSDKINITKQITTKKGIDGE 1020
QY 1021 DSSSDATSNANLTIKTRELKLTEDLSISGFNKAEITAKDGRDLTIGNSDGNSGAEAKTV 1080
DB 1021 DSSSDATSNANLTIKTRELKLTEDLSISGFNKAEITAKDGRDLTIGNSDGNSGAEAKTV 1080
QY 1081 TFNNVKDSKISADGHNVNLTNSKVTSSSNGGRESNSDNDTGLTITAKNVENVKDIITSLKT 1140
DB 1081 TFNNVKDSKISADGHNVNLTNSKVTSSSNGGRESNSDNDTGLTITAKNVENVKDIITSLKT 1140
QY 1141 VNITASEKVTTAGSTINATNGKASITTKTGDIISGTSIGNTVSVSATVDLTITKSGSKIEA 1200
DB 1141 VNITASEKVTTAGSTINATNGKASITTKTGDIISGTSIGNTVSVSATVDLTITKSGSKIEA 1200
QY 1201 KSGEANTVSATGTIGTISGTVNVTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
DB 1201 KSGEANTVSATGTIGTISGTVNVTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
QY 1261 GSSITTKGVDLLAONGSTAGSINAANVTLTNTGTLTTVAGSDIKATSGTLVINAKDAK 1320
DB 1261 GSSITTKGVDLLAONGSTAGSINAANVTLTNTGTLTTVAGSDIKATSGTLVINAKDAK 1320
QY 1321 LMGDASGDSDEVNAVNASGSGSVTAATSSVNITGDLNVTGNGLIISKDGRNTVRLRGKE 1380
DB 1321 LMGDASGDSDEVNAVNASGSGSVTAATSSVNITGDLNVTGNGLIISKDGRNTVRLRGKE 1380
QY 1381 IEVKYIQPGVASVEEVIEAKRVLEKVDLSDEERETLAKLGVSAVRFVPEPNNITVNTQN 1440
DB 1381 IEVKYIQPGVASVEEVIEAKRVLEKVDLSDEERETLAKLGVSAVRFVPEPNNITVNTQN 1440
QY 1441 EFTTRPSSQVILISEGKACFSSGNGARVCTNVAADDGQP 1477
DB 1441 EFTTRPSSQVILISEGKACFSSGNGARVCTNVAADDGQP 1477

RESULT 6

US-08-617-697-4
Sequence 4, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

us-10-092-880-4.rai

Tue Mar 25 08:37:45 2003

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-4

Query Match 100.0%; Score 7407; DB 2; Length 1477;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLISLGV 60
Db 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLISLGV 60
Qy 61 SIPOSVLASGLQGVVHGTATMGVGNKTIIRNSVDIAIINWKOFNDQEMVQFOLENN 120
Db 61 SIPOSVLASGLQGVVHGTATMGVGNKTIIRNSVDIAIINWKOFNDQEMVQFOLENN 120
Qy 121 NSAVFNRYTSNQISOLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDSNENIK 180
Db 121 NSAVFNRYTSNQISOLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDSNENIK 180
Qy 181 ARNETFEQTKDKALAEIIVHGLITVKGDSVNLIGGKVKNEGVSIVNGGSSISLLAGOKIT 240
Db 181 ARNETFEQTKDKALAEIIVHGLITVKGDSVNLIGGKVKNEGVSIVNGGSSISLLAGOKIT 240
Qy 241 ISDIINPTITYIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSVDKSGNIV 300
Db 241 ISDIINPTITYIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSVDKSGNIV 300
Qy 301 LSARGEAEFISQAQNOQAKGGLMITGDKVTLTKGAVIDLKSGEGGTYLGGDERGE 360
Db 301 LSARGEAEFISQAQNOQAKGGLMITGDKVTLTKGAVIDLKSGEGGTYLGGDERGE 360
Qy 361 GRNGIOLAKKTSLEKGSITNSGKEGFAIVWGDIALIDGNINAQSGDIAKTGGFVET 420
Db 361 GRNGIOLAKKTSLEKGSITNSGKEGFAIVWGDIALIDGNINAQSGDIAKTGGFVET 420
Qy 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGINDPEPTGTGEASDPKNSL 480
Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGINDPEPTGTGEASDPKNSL 480
Qy 481 TLTNTTISNYLKNWMTNMTASRLITVNSSINIGSNHILHSGKRGQGGVQIDGDITS 540
Db 481 TLTNTTISNYLKNWMTNMTASRLITVNSSINIGSNHILHSGKRGQGGVQIDGDITS 540
Qy 541 KGNLTIYSGGWVVDVHKNITLQDGFNLITAAVAFEGGNKARDAANAKIVAQGTVTITG 600
Db 541 KGNLTIYSGGWVVDVHKNITLQDGFNLITAAVAFEGGNKARDAANAKIVAQGTVTITG 600
Qy 601 EGKDFRANVSLNGCTCKGLNISSVNNLTHNLSGTINISGNITINOTTRKNTSYWQTS 660
Db 601 EGKDFRANVSLNGCTCKGLNISSVNNLTHNLSGTINISGNITINOTTRKNTSYWQTS 660

; Sequence 4, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.

```



```

; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-4

```

```

Query Match      100.0%; Score 7407; DB 4; Length 1477;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGESEKPKARKMVRHLALPLSAMLISLQVT 60
DB 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGESEKPKARKMVRHLALPLSAMLISLQVT 60

QY 61 SIPOSVLASGLQGMVDVHGATMDVGNKTIIRNSVDIAIINWKQFIDQNEWQFLOENN 120
DB 61 SIPOSVLASGLQGMVDVHGATMDVGNKTIIRNSVDIAIINWKQFIDQNEWQFLOENN 120

QY 121 NSAVENRVTNSQISQLKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDIISNENIK 180
DB 121 NSAVENRVTNSQISQLKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDIISNENIK 180

QY 181 ARNFTFEQTKKALAEIVNHLITVKGDSYNLIGGKVKNEGVISVNGGSIISLLAGOKIT 240
DB 181 ARNFTFEQTKKALAEIVNHLITVKGDSYNLIGGKVKNEGVISVNGGSIISLLAGOKIT 240

QY 241 ISDIINPITTSIAAPENEAVALNIGDIFAKGGINVRAATIRNQKLSADSVDKSGNIV 300
DB 241 ISDIINPITTSIAAPENEAVALNIGDIFAKGGINVRAATIRNQKLSADSVDKSGNIV 300

QY 301 LSAKEGEAEIGGVISAQOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360
DB 301 LSAKEGEAEIGGVISAQOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360

QY 361 CKNGIOLAKKTSLEKGSTINVSKEGKGFAIIVWGDIALIDGNINAQSGDIAGTGGFVET 420
DB 361 CKNGIOLAKKTSLEKGSTINVSKEGKGFAIIVWGDIALIDGNINAQSGDIAGTGGFVET 420

QY 421 SGHDLFIKDAIVAIDAKEWLLDFDNVSNIAEDPLFNNTGINDERPTGTGEASDPKKNSELK 480
DB 421 SGHDLFIKDAIVAIDAKEWLLDFDNVSNIAEDPLFNNTGINDERPTGTGEASDPKKNSELK 480

QY 481 TTLNTTISNVLKNAWTMTNITASRKLTVNSSINIGNSNHLILHSKQGGGVQIDGDIITS 540
DB 481 TTLNTTISNVLKNAWTMTNITASRKLTVNSSINIGNSNHLILHSKQGGGVQIDGDIITS 540

```

```

QY 541 KGNLTIYSGGWVDVHKNTLDQGFNLITAAVAFEGGNKARDAAANAKIVAQGVITIG 600
DB 541 KGNLTIYSGGWVDVHKNTLDQGFNLITAAVAFEGGNKARDAAANAKIVAQGVITIG 600

QY 601 EGDFRANNVSLNGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTRKNTSYQWTSHD 660
DB 601 EGDFRANNVSLNGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTRKNTSYQWTSHD 660

QY 661 SHWNVSALENGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTRKNTSYQWTSHD 720
DB 661 SHWNVSALENGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTRKNTSYQWTSHD 720

QY 721 KLPKNENMTSKPLPIRPLANITATGGGVFFDIYANISGRGAELKMEISEINSGANFTL 780
DB 721 KLPKNENMTSKPLPIRPLANITATGGGVFFDIYANISGRGAELKMEISEINSGANFTL 780

QY 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFDYDGVARNAINSTYISILGNNVTIG 840
DB 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFDYDGVARNAINSTYISILGNNVTIG 840

QY 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDVIRKGLSLLVNGSLTGENADIKGN 900
DB 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDVIRKGLSLLVNGSLTGENADIKGN 900

QY 901 LTISESATFKGKTRDTLNTGNFTNNGTAEINITQGVVVKLGNVTNDGDLNITTHAKRNQ 960
DB 901 LTISESATFKGKTRDTLNTGNFTNNGTAEINITQGVVVKLGNVTNDGDLNITTHAKRNQ 960

QY 961 SIIGDIIINKKGLNITDSNNDAEIQIGGNIQKGEGLNITISSDKINITKQITIKKGIDG 1020
DB 961 SIIGDIIINKKGLNITDSNNDAEIQIGGNIQKGEGLNITISSDKINITKQITIKKGIDG 1020

QY 1021 DSSSDATSNANLTIKTRELKLTEDLSISGFNKAETITAKDGRDLTIGNSDNGSGAEAKTV 1080
DB 1021 DSSSDATSNANLTIKTRELKLTEDLSISGFNKAETITAKDGRDLTIGNSDNGSGAEAKTV 1080

QY 1081 TFNNVKOSKISADGHNVTLNSKVKTSNNGSGRESNSDNDTGLTITAKNVEVNDITSLKT 1140
DB 1081 TFNNVKOSKISADGHNVTLNSKVKTSNNGSGRESNSDNDTGLTITAKNVEVNDITSLKT 1140

QY 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEA 1200
DB 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEA 1200

QY 1201 KSGEANVTSAITGTIGGTISGNTVNTANAGDLTVNGCAEINATEGAATLTATGNTLTTEA 1260
DB 1201 KSGEANVTSAITGTIGGTISGNTVNTANAGDLTVNGCAEINATEGAATLTATGNTLTTEA 1260

QY 1261 GSSITSTKGOVDLLAQNAGSTAGSINAAVNTLNTTGLTTTVAGSDIKATSGTLVINAKDAK 1320
DB 1261 GSSITSTKGOVDLLAQNAGSTAGSINAAVNTLNTTGLTTTVAGSDIKATSGTLVINAKDAK 1320

QY 1321 LNGDASGDSPEVNAVNASGSGSVTAATSSSVNITGDLTVNGLNIIISKDGRNTRVRGKE 1380
DB 1321 LNGDASGDSPEVNAVNASGSGSVTAATSSSVNITGDLTVNGLNIIISKDGRNTRVRGKE 1380

QY 1381 IEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLGVSAVRFVBPNNITVNTON 1440
DB 1381 IEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLGVSAVRFVBPNNITVNTON 1440

QY 1441 EFTTRPSSQVILISEGKACFSNGARVCTNVADDDQOP 1477
DB 1441 EFTTRPSSQVILISEGKACFSNGARVCTNVADDDQOP 1477

```

```

RESULT 8
US-09-206-942-71
; Sequence 71, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping

```

```

; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: Molecular Weight Proteins
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 71
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-206-942-71

Query Match      98.0%; Score 7261; DB 4; Length 1477;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1454; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY 1 MNKIYRLKFKSRNALVANVELARGCDHSTKSEKPAKVPKHLALKPLSAMLISLGVY 60
DB 1 MNKIYRLKFKSRNALVANVELARGCDHSTKSEKFRVYTFRCNHLALKPLSAMLISLGVY 60
QY 61 SIPOSVLASLGQMDVVGHTATVQVGNKTIIRNSVDAILNKKQFNDONEMVQFOLENN 120
DB 61 SIPOSVLASLGQMDVVGHTATVQVGNKTIIRNSVDAILNKKQFNDONEMVQFOLENN 120
QY 121 NSAVFNRTVNSOISQKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
DB 121 NSAVFNRTVNSOISQKGLDSNGOVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
QY 181 ARNFTFQTKDKALAEIVNHLITVKGDSVNLIGGKRVKNEGVIYVNGGSIISLLAGOKIT 240
DB 181 ARNFTFQTKDKALAEIVNHLITVKGDSVNLIGGKRVKNEGVIYVNGGSIISLLAGOKIT 240
QY 241 ISDIINPTIYSAAPENEAVALNLDIFAKGNNINVRATIRNOGKLSADSVSKDKSGNIV 300
DB 241 ISDIINPTIYSAAPENEAVALNLDIFAKGNNINVRATIRNOGKLSADSVSKDKSGNIV 300
QY 301 LSAKEGEAEIGGVISAQNOQAQKGLMTGDKVTGKGTGAVIDLSGKEGTYLGGDERGE 360
DB 301 LSAKEGEAEIGGVISAQNOQAQKGLMTGDKVTGKGTGAVIDLSGKEGTYLGGDERGE 360
QY 361 KNGIOLAKKTSLEKSTINVSQKKGFPATVWGDIADIGNINAQSGGDIAGTGPFVET 420
DB 361 KNGIOLAKKTSLEKSTINVSQKKGFPATVWGDIADIGNINAQSGGDIAGTGPFVET 420
QY 421 SGHDLFTKDAIVDAKEMLLDFDWNSTNAEDPLENNGTGDEPTGTGEASDPKKNSELK 480
DB 421 SGHDLFTKDAIVDAKEMLLDFDWNSTNAEDPLENNGTGDEPTGTGEASDPKKNSELK 480
QY 481 TTLNTTISNLYLKNAWTMNITASRLKLVNSSINTGNSHLLHSKGQGGGVQIDGDIITS 540
DB 481 TTLNTTISNLYLKNAWTMNITASRLKLVNSSINTGNSHLLHSKGQGGGVQIDGDIITS 540
QY 541 KGGNLTYSGGWVDVHKNTIDQGFNLITAAVAFEGGNKARDAANAIAVQGTVTITG 600
DB 541 KGGNLTYSGGWVDVHKNTIDQGFNLITAAVAFEGGNKARDAANAIAVQGTVTITG 600
QY 601 EGKDFRANNSVLTGKGLNISSVNNLTNLSGTINTSGNITINOTTRKNTSYWQTSHD 660
DB 601 EGKDFRANNSVLTGKGLNISSVNNLTNLSGTINTSGNITINOTTRKNTSYWQTSHD 660
QY 661 SHWNYSALNLEGTGANFTFIKVISSNSKLTTOYRBSAGVNFNGVNGNSFNLKGAQVNF 720
DB 661 SHWNYSALNLEGTGANFTFIKVISSNSKLTTOYRBSAGVNFNGVNGNSFNLKGAQVNF 720
QY 721 KUKPENENNTSKPLDIRFLANITATGGGVSFFDIYANHSRGAEKUKSEINISGANFTL 780
DB 721 KUKPENENNTSKPLDIRFLANITATGGGVSFFDIYANHSRGAEKUKSEINISGANFTL 780
QY 781 NSHVRGDDAFKINKDLTINATNSNFSLRQTKDDFYDGYARNAINSTYNISILGGNVTLGG 840

```

RESULT 9

```

US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 71.3%; Score 5282; DB 2; Length 1600;
Best Local Similarity 68.3%; Pred. No. 1.1e-292;
Matches 1099; Conservative 136; Mismatches 231; Indels 142; Gaps 16;

QY 1 MNKIYLFKSKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALPLSALLSLGVT 60
DB 1 MNKIYLFKSKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALPLSALLSLGMA 60

QY 61 SIPOSVLASGLQGMVDVHGTATQMDVGNKTIIRNSVDIAIINNKQFNIDQNMVQFLOENN 120
DB 61 SIPOSVLASGLQGMVDVHGTATQMDVGNKTIIRNSVDIAIINNKQFNIDQNMVQFLOESS 120

QY 121 NSAVENRVTNQISQKLGILDSNGOVFLINPNGITIGKDAIINNGTFASTLDISNENIK 180
DB 121 NSAVENRVTNQISQKLGILDSNGOVFLINPNGITIGKDAIINNGTFASTLDISNENIK 180

QY 181 ARNFFTEQTKDALKAEIYVHGLITVGDGYNLIGGVKNKGVSIVNGGSIISLLAGOKIT 240
DB 181 ARNFFTEQTKDALKAEIYVHGLITVGDGYNLIGGVKNKGVSIVNGGSIISLLAGOKIT 240

QY 241 ISDIINPTITYSIAAPENAVNLGDIKFAKGNINVRATIRNQQKLSADSVSKDKSGNIV 300
DB 241 ISDIINPTITYSIAAPENAVNLGDIKFAKGNINVRATIRNQQKLSADSVSKDKSGNIV 300

QY 301 LSAKEGEAEIGGVISAQNOQAQKGLMITGDKVTLTKTGAVIDLSGKGGETYLGGERGE 360
DB 301 LSAKEGEAEIGGVISAQNOQAQKGLMITGDKVTLTKTGAVIDLSGKGGETYLGGERGE 360

QY 361 GKNGIOLAKTSLBKGSTINVSKEGKGFAIVWGDIALIDGNINAQSGDIAGTGGFVET 420
DB 361 GKNGIOLAKTSLBKGSTINVSKEGKGFAIVWGDIALIDGNINAQSGDIAGTGGFVET 419

QY 421 SCHDLFTKDIAVDAKELWLLDFONVSNIAEDPLFNNTGIDNDEPTGTGEASDPKPSKSELK 480
DB 420 SCHDLSTGDVDVDAKELWLLDPDDVSTETTSGRNNTGENQGYTGDGTGKSPKGSNISK 479

QY 481 TTLNTNTISYLNKAWMTNITASKRLTVNSINIGSNHLSHLHSKQGGGVQIDGDTIS 540
DB 480 PTLTNTLEQLRSGSVNITANNRIYVNSINL-SNGSLTLTKRD--GVKINGDITS 535

QY 541 -KGNLTIYSGWVDVHKNITLQOGLFNITAA-SVAFEGGNKRNKARDAANAIVAOQTIVRI 598
DB 536 NENGLNTIKAGSWVDVHKNITLQOGLFNITAA-SVAFEGGNKRNKARDAANAIVAOQTIVRI 595

QY 599 TGECKDFRANVSLNGTGGKGLNLTISVSNLNLSTINISGNTIINQITRNKTSWQTS 658
DB 596 NKDDQFRFNVSNGTGGKGLNLTISVSNLNLSTINISGNTIINQITRNKTSWQTS 655

QY 659 HDGHNVSALNLETGANFTFIKYISSNKG--LTTQYRSSAGVNFNGVNGNMSFNLKEGA 716
DB 656 KDSYWNVSSTLTNTVQKFTFIKFDVSGSNGQDLRRSRRSFAGVHFNIGKTNFNIGANA 715

```

```

QY 717 KVNEKLKPNENMNTSKPLRFLANITATGG--GSVFDDIYANHSGRAEAEUKMSEINISN 774
DB 716 KALFKALPNAADPKKELPITFNFANITATGNSDSSVAFDIIHANLITSAAGINMDSINITG 775
QY 775 GANFTLNSHVRGDDAFKINKDLTINATNSFSLRQTKDDFDYGVARNAINSTYNIISLGG 834
DB 776 GLDFSITSHNRNSNAFEIKKDLTINATGNSFSLRQTKDDFDYGVARNAINSHNLITLGG 835
QY 835 NVTLGGONSSSITGNITIEKAAVTLLEANNAPQOQNIROIRVIKLGLSLLVNGSLSLTGEN 894
DB 836 NVTLGGONSSSITGNITINKANVTLOADTNSNTGLKRTLTPLGINSVEGNSLTLTGAN 895
QY 895 ADIKGNLTISAEATEKGTTRDTLNTCNFTNNGTAEINITQGVVYKLVNVDGDLNITTH 954
DB 896 ANIVGNLSIAEDSTFKGEASDNLTITFTNNGTANINI-KGVVYKLVNVDGDLNITTH 954
QY 955 AKRNQRIIGDIIINKGSLNITDSNNDAEIQIGGINSQKQEGNLTSSDKINITKQITIK 1014
DB 955 ASGTQKTIINGNITNEKGLNINIKAKADAEIQIGGINSQKQEGNLTSSDKINITKQITIK 1014
QY 1015 KGIDGEDSSSDATSNANLTITKTELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGNSG 1074
DB 1015 AGVEGGRSDSEAEANLTIQTKELKLAGLNISGFNKAEITAKGSDLTIGNASGNN-- 1072
QY 1075 AEAKTVTFNNVKDSKISADGHNVTLSKVKVTSNNGSGRESNSNDTGLTITAKNVEVNDK 1134
DB 1073 ADARKVTDFKVKDSKISTDGHNVTLNSEVKT--SNSSNAGNDSNGLTISAKDVTNNN 1130
QY 1135 ITSUKTVNI-----TASEKVTIT 1152
DB 1131 VTSKHTINISAAAGNVTTKEGTINATGTSVEVTAQNGTIKGNITSQNVTVTATENLVTT 1190
QY 1153 AGSTINATNGKASITTKTGDISG-----TISGNTVSVS-----ATVD---L 1190
DB 1191 ENAVINATSGTVNISVTKGDIKGIEBTSQNVNITAGMTLVKVSNTGQDVTVTADAGAL 1250
QY 1191 TTKSGSKIEAKSEANVTSATG-----TISGNTVSVS-----TASEKVTIT 1212
DB 1251 TTTAGTISATGNNANITTKTGDISGKVSSESVTLVATGATLAVGNISGNTVTITADS 1310
QY 1213 -----TIG-----GTISGNTVNTANAGDLTVGNAEINATEGAAT 1248
DB 1311 GKLTISVGTINGTNSVTTSSSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAAT 1370
QY 1249 LRTATGNLTTEAGSSITSTKGQVDDLLAQNAGSIAGSINAANVTNTTGLTITVAGSDIKAT 1308
DB 1371 LTAESKLTQTGSSITSSNGQTTLTAKDSSIAGNINAANVTNTTGLTITVAGSDIKAT 1430
QY 1309 SGTFLVINAKDLNGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVNGLNITISK 1368
DB 1431 SGTFLVINAKDLNGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVNGLNITISK 1490
QY 1369 DGRNTVRLRGKEIEVKYIQPGVASVEEVEAKRVLEKVKDLSDERETLAKLGVSAVRPV 1428
DB 1491 NGRNTVRLRGKEIEVKYIQPGVASVEEVEAKRVLEKVKDLSDERETLAKLGVSAVRPV 1550
QY 1429 EPNNITVNTQNEFTTRPSSQVLIISGKACFSNGARVCTNVAADDGQ 1476
DB 1551 EPNNITVNTQNEFTTRPSSQVLIISGKACFSNGARVCTNVAADDGQ 1598

```

RESULT 10

```

US-09-206-942-73
; Sequence 73, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb

```

Tue Mar 25 08:37:45 2003

; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 73
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-73

Query Match 69.7%; Score 5164; DB 4; Length 1036;
Best Local Similarity 99.5%; Pred. No. 3.4e-286;
Matches 1030; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 443 DNVISINAEPLFNNTGINDPEFTGTGEASDPKKNSELKTTLTNTTISNYLKNAMTNITA 502
DB 2 DQVTEADPLRNNTGINDPEFTGTGEASDPKKNSELKTTLTNTTISNYLKNAMTNITA 61
QY 503 SRKLTWNSSINIGNSHLILHSKGGGQVQIDGDIITSGGGLTIYSGGWVDVHKNTILD 562
DB 62 SRKLTWNSSINIGNSHLILHSKGGGQVQIDGDIITSGGGLTIYSGGWVDVHKNTILD 121
QY 563 QGFLNTAASVAFEGGNKARDAANAKIVAQGTVTITCEGKDFRANNVSLNCTGKGLNII 622
DB 122 QGFLNTAASVAFEGGNKARDAANAKIVAQGTVTITCEGKDFRANNVSLNCTGKGLNII 181
QY 623 SSYNNLTHNLSGTINISGNITINOTTRKNTSYWQTSHDHSHWVSALNLETGANFTFIKYI 682
DB 182 SSYNNLTHNLSGTINISGNITINOTTRKNTSYWQTSHDHSHWVSALNLETGANFTFIKYI 241
QY 683 SSNSKGLTTQYRSSAGVNFNGVNGNSFNLKEGAKVFNKLPKNENMNTSKPLPIRELANI 742
DB 242 SSNSKGLTTQYRSSAGVNFNGVNGNSFNLKEGAKVFNKLPKNENMNTSKPLPIRELANI 301
QY 743 TATGGGSVFDDIYANHSRGAELKMEIINISGANFTLNSHVRGDDAFKINKDLTINATN 802
DB 302 TATGGGSVFDDIYANHSRGAELKMEIINISGANFTLNSHVRGDDAFKINKDLTINATN 361
QY 803 SNFSLRQTKDDFYDYGARNAINSTYINISILGNGVTILGGNSSSITGNITIEKAANVTLE 862
DB 362 SNFSLRQTKDDFYDYGARNAINSTYINISILGNGVTILGGNSSSITGNITIEKAANVTLE 421
QY 863 ANNAPNOONTRDRVILKGLSVNGSLTGENADIKGNLTISESATFKGTRDTLTNITGN 922
DB 422 ANNAPNOONTRDRVILKGLSVNGSLTGENADIKGNLTISESATFKGTRDTLTNITGN 481
QY 923 FTNNGTAEINITOGVYKLGWNTDGLNITTHAKRNORSITIGDIIINKKGSNLITDSND 982
DB 482 FTNNGTAEINITOGVYKLGWNTDGLNITTHAKRNORSITIGDIIINKKGSNLITDSND 541
QY 983 AETIOIGNISOKENLTISDKININIKQITIKKIDGEDSSDATSNANLTITKTELKLT 1042
DB 542 AETIOIGNISOKENLTISDKININIKQITIKKIDGEDSSDATSNANLTITKTELKLT 601
QY 1043 EDLSISGFENKAEITAKDGRDLTIGNSDNGSGAPAKTVTFNNVKDSKISADGHNVTLNSK 1102
DB 602 EDLSISGFENKAEITAKDGRDLTIGNSDNGSGAPAKTVTFNNVKDSKISADGHNVTLNSK 661
QY 1103 VKTSSNGGRESNDNDTGLTITAKNVEVNRKIDITSLKTVNITASEKVTITTAGSTINATNG 1162
DB 662 VKTSSNGGRESNDNDTGLTITAKNVEVNRKIDITSLKTVNITASEKVTITTAGSTINATNG 721
QY 1163 KASITTKTIGDLSGTISGNTVSATVDLTITKSGSKI EAKSGEANVT SATCTIGTISGNT 1222
DB 722 KASITTKTIGDLSGTISGNTVSATVDLTITKSGSKI EAKSGEANVT SATCTIGTISGNT 781
QY 1223 NVVTANAGDLTVNGCAELNATEGAATLTATCNTLTITTAGSSITSTKGOVDLLAONGSIAG 1282
DB 782 NVVTANAGDLTVNGCAELNATEGAATLTATCNTLTITTAGSSITSTKGOVDLLAONGSIAG 841
QY 1283 SINAANVTLNTTGTTLTVAGSDIKATSGTLVINAKDAKLGNDASGDSTEVNAVNASGSGS 1342

DB 842 SINAANVTLNTTGTTLTVAGSDIKATSGTLVINAKDAKLGNDASGDSTEVNAVNASGSGS 901
QY 1343 VTAATSSVNITGDLNTVNGLNIIISKDGRNTRVLRGKEIEVKYIQPGVASVEVIEAKRV 1402
DB 902 VTAATSSVNITGDLNTVNGLNIIISKDGRNTRVLRGKEIEVKYIQPGVASVEVIEAKRV 961
QY 1403 LEKVKDLSDDEERETLAKLIGSAVRFEVPEPNTITVNTQNEFTTRPSSQVITISEGKACFSSG 1462
DB 962 LEKVKDLSDDEERETLAKLIGSAVRFEVPEPNTITVNTQNEFTTRPSSQVITISEGKACFSSG 1021
QY 1463 NGARVCTNVADDGOP 1477
DB 1022 NGARVCTNVADDGOP 1036
RESULT 11
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-682-2
Query Match 67.1%; Score 4967; DB 1; Length 1536;
Best Local Similarity 67.3%; Pred. No. 9.3e-275;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRLHAKLPKLSAMLLSLGVT 60
DB 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRLHAKLPKLSAMLLSLGVT 60
QY 61 SIPQSVLASGLOGMDVYVHGVTATMQVDGNKTIIRNSVDAIINNKQPNIDONEMVQFLQENN 120
DB 61 SIPQSVLASGLOGMDVYVHGVTATMQVDGNKTIIRNSVDAIINNKQPNIDONEMVQFLQENN 120
QY 121 NSAVFNRVTSNQTISQLKILDSNGQVFLINPNGITIGKDAINTNGFTASTLIDISNETIK 180
DB 121 NSAVFNRVTSNQTISQLKILDSNGQVFLINPNGITIGKDAINTNGFTASTLIDISNETIK 180

Db 121 NSAVFNRTSNQISOLKGLSDNSQGVFLINPNQITIGKDAIINTNGFTASTLDISNENIK 180
QY 181 ARNFTPEQTKD KALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGOKIT 240
Db 181 ARNFTPEQTKD KALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGOKIT 240
QY 241 ISDIINPTIYSTAAPENAVNLGDIFAKGGINIVRAATIRNOGKLSADSVSKDGSNIV 300
Db 241 ISDIINPTIYSTAAPENAVNLGDIFAKGGINIVRAATIRNOGKLSADSVSKDGSNIV 300
QY 301 LSAKEGEAIEGIVISAQNOQAAGKGLMITGDKVTLKTGAVIDLSGKEGETVLGDERGE 360
Db 301 LSAKEGEAIEGIVISAQNOQAAGKGLMITGDKVTLKTGAVIDLSGKEGETVLGDERGE 360
QY 361 GKGIOAKKTSLEKGSTINVSKEKGGPAIWMGDIALIDGNNIAQSGDIAKTGGFVET 420
Db 361 GKGIOAKKTSLEKGSTINVSKEKGGPAIWMGDIALIDGNNIAQSGDIAKTGGFVET 420
QY 421 SGHDLFIKDAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479
Db 421 SGHDLFIKDAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479
QY 480 KTLTNTTISNVLKNAWNTNITASRKLIVNNSINIGSNSHLILHSGQGGVQIDGDIT 539
Db 479 KTLTNTTISNVLKNAWNTNITANORIVYVNSSINL-SNGSLTLWSEGRSGGVEINNDIT 537
QY 540 ----SKGGLTIYSGGWVDVHKNIITLD--QGFLNITA-ASVAFEGGNKARDAANAKIVAQ 593
Db 538 TGDDTRGANLTIYSGGWVDVHKNIITLD--QGFLNITA-ASVAFEGGNKARDAANAKIVAQ 593
QY 594 GTVITGEGKDFRANVSLNLTGKGLNISSVNN---LTHNSLTINISGNITINOTTRK 650
Db 591 GTIT--SGNQKGRFRNVLNLTGSGLOQTKRTNKYAITNFEGLNLSGKVNLSMVLPK 649
QY 651 NTS-YWOTSHDSHWNYSALNLETGANETTF-LKYISSNSKGLITQYRSAGVNFVNGVNM 708
Db 650 NESGVDKFKGTYNWLNSLWSEGEFNLITDSRGSDSAGTLTQPYNLNGISF---NKDT 706
QY 709 SPNUKEGAKVNFKLKPNENMNTSKPLP-IRFLANITATGGGVSFFDIYANHS---RGAE 764
Db 707 TFNVARNARVDFIKAPIGINKYSLNYSVNGNISVSGGSDVFTLLASSNVQTPGVV 766
QY 765 LKMEISINSGANFTLNSHVRGDDAFKINKDLTINATNSNFSRQTKDDPYDGYARNAIN 824
Db 767 INSKYFNVSTGSSRFTSGTKTGFSLKDLTLNATGNGNITLQVEGT--DGMIGKGIV 824
QY 825 STYNISLGGNVTLGGQSSSITGNITIEKAANVTLEANNAPNOQNRDRVTKLGLSLV 884
Db 825 AKKNITFEGGNITFGSRKAVTEIEGNVTINNANVTLLIGSDFDNHQ--KPLTIKDVIIIN 882
QY 885 NGSLSLGENADIKGLNLTISESATFGKTRDTLNTGNTNNGTAEINITQGVYKLGNYT 944
Db 883 SGNLTAGNVNITAGNLTIVESNANFAITNFTFNVGGLFDNKGNSNISIAKGGAREFKDID 942
QY 945 NQDGLNITTHAKNRQSIIGDIIINKGSLNITSDNSNDAEIQIGGNSIOKEGNLTISSDK 1004
Db 943 NSKNLSITTSSTYRTIISGNTNKGDLNITNEDSGDEMQIGGVDVSQKEGNLTISSDK 1002
QY 1005 INTKQITIKGIDGEDSSDATSNANLTIKTKELKLTDLTISGFNKAETIAKGRDIT 1064
Db 1003 INTKQITIKAGVDGSDSDATSNANLTIKTKELKLTDLTISGFNKAETIAKGRDIT 1062
QY 1065 IGNSDNGSCAEAKTVFNVRKDSKISADGHVNTLNSKVTKSSNGGRESNDNDTGLTI 1124
Db 1063 IGNTNSAD-GTNAKKVTFFNOVKDSKISADGHVNTLNSKVTKSSNGGRESNDNDTGLTI 1121
QY 1125 TAKNVNNDITSLKTVNITA-SEKVTITAGSTINATNKAETIAKGRDIT 1170
Db 1122 DAKNVVNNITSHKAVSISATSGEITTKGTGTINATGNVEITAQTSILGIESSSGS 1181
QY 1171 -----GDISGTISGNT 1181
Db 1182 VTLTATEGALAVSNISGNTVTVTANSALTTLIAGSTIKGTESVTTSSQSGDIGGTISGNT 1241

QY 1182 VSVSATVOLTTSKSGKIEAKSGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEIN 1241
Db 1242 VEVKATLTITQSSNKKIKATTEANVTISATGTIGGTISGNTVNTANAGDLTVNGAEIN 1301
QY 1242 ATEGAATLTATGNTLTTEAGSSITSTKQVOLLAQNGSIAGSINAANVTLLMTTGLTVA 1301
Db 1302 ATEGAATLTITSSGKLTTEASSHITSAGQVNLSDAQDSVAGSINAANVTLLMTTGLTVA 1361
QY 1302 GSDIKATSGTLVINAADKALMGDASGDSTEVNANVAGSGSVTAATSSSVNITGDLTAVN 1361
Db 1362 GSNINATSGTLVINAADKALMGDASGDSTEVNANVAGSGSVTAATSSSVNITGDLTAVN 1421
QY 1362 GLNIIISKDRNTVPLRKEIEVKYIQGVASVEVIEAKRVLEKVKDLSDEERETLAKLG 1421
Db 1422 GLNIIISKDRNTVPLRKEIEVKYIQGVASVEVIEAKRVLEKVKDLSDEERETLAKLG 1481
QY 1422 VSAVRFEPPNNTITVNTQNEFTTTPSSQVITSECKACFSSGSGARVCTNVADDOG 1476
Db 1482 VSAVRFEPPNNTITVNTQNEFTTTPSSQVITSECKACFSSGSGARVCTNVADDOG 1536

RESULT 12

US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302.832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 67.1%; Score 4967; DB 1; Length 1536;
Best Local Similarity 67.3%; Pred. No. 9.3e-275;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

QY 1 MNKYRLKFKSRNALVAVSELARGCDHSTKEGSEKPARMKVRLHAKLPKLSAMLLSLGVT 60
Db 1 MNKYRLKFKSRNALVAVSELARGCDHSTKEGSEKPARMKVRLHAKLPKLSAMLLSLGVT 60
QY 61 SIQSVLASGLQGMVHVHGTATMOVDGKNTIIRNSVDIAIINNKQFIDQNMVQFLOQNN 120
Db 61 SIQSVLASGLQGMVHVHGTATMOVDGKNTIIRNSVDIAIINNKQFIDQNMVQFLOQNN 120
QY 121 NSAVFNRVTNOISOLKGLDNGOVFLINPNGITIGTKDAIINTNGFTASTLDSINENIK 180
Db 121 NSAVFNRVTNOISOLKGLDNGOVFLINPNGITIGTKDAIINTNGFTASTLDSINENIK 180
QY 181 ARNFTFQTKDKALAEIVNHGLITVKGDSVNLGGKVEGVISVNGGSIISLLAGOKIT 240
Db 181 ARNFTFQTKDKALAEIVNHGLITVKGDSVNLGGKVEGVISVNGGSIISLLAGOKIT 240
QY 241 ISDIINPTIYTAAPNEAVNLGDIIPAKGGINVRAATIRNOGKLSADSVSKDGSNTV 300
Db 241 ISDIINPTIYTAAPNEAVNLGDIIPAKGGINVRAATIRNOGKLSADSVSKDGSNTV 300
QY 301 LSAKEGEAEIGGVISAQNOQAKGKMLITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360
Db 301 LSAKEGEAEIGGVISAQNOQAKGKMLITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360
QY 361 KGNGIOLAKKTSLEKGGSTINVSKEGKGFATVWGDIADIDGNINAQSGDIAKTGGFVET 420
Db 361 KGNGIOLAKKTSLEKGGSTINVSKEGKGFATVWGDIADIDGNINAQSGDIAKTGGFVET 420
QY 421 SGHDLFIKDNAIVDAKEMWLLDFDNVSIINAEPLFNNTGGINDEFTGTG-EASDPKKNSEL 479
Db 421 SGHDLFIKDNAIVDAKEMWLLDFDNVSIINAEPLFNNTGGINDEFTGTG-EASDPKKNSEL 479
QY 480 KTTLTNTTISVNLKAWMTNMTASRKLTVNSSINIGSNHLSHLSKQGGVQIDGDIIT 539
Db 479 KTTLTNTTLESLLKKGTFVNITANQRIYVNSSINL-SNGSLTWSRSGGGVEINNDIT 537
QY 540 ---SKGGLNLTISGSGWVHKNITLD-QGFLNITA-ASVAFEGGNKARDAAAKIAVQ 593
Db 538 TGDTRGANLTISGSGWVHKNISLGAQGNINITAKODIAFEKSGNOV-----ITQ 590
QY 594 GTVITTEGKDFRANVSLNKGKLNLSVNN--LTHNLSTINTINSGITINOTRK 650
Db 591 GTIT-SGNQGRFRNVSNGTSGGLOFTTKTKYATNFKEGTLNLSGKVNISMVLPK 649
QY 651 NTS-YWQTSHDHWNVALNLETGANFTF-IKYSNSKGLTQYRSAGVNFVNGVNM 708
Db 650 NESGYDKFKRTYWNLTSLNVSSEGEFNLIDSRGSDSAGTLTQYNLNGISF---NKDT 706
QY 709 SFNLKGAQVNFKLKPNENMNTSKPLP-IRFLANITATGGGVFFDIYANHSG---RGAE 764
Db 707 TFNVNARNVDFDIKAPIGINKYSSLVNYSFNGNITSVSGGSDVDFTLASSNNVOTPGV 766
QY 765 LKMSINISGANFTLNSHVGRDDAFKINKOLITINATSNFSLRQTKDDFYDGYARNAIN 824
Db 767 INSKYFNVSTGSLRFTKSGTKTGFSEIKDLTLNATGNITLLQVEGT--DGMITGKGV 824
QY 825 STYNISILGNVTGSGONSSSITIGNITIEKAANTLEANNAPNOQNRDRVILKGLSLLV 884
Db 825 AKKNITFEKNITFGSRKAVTELEGNVTINNNANVTLLGSDFDNHQ--KPLTIKKDVAIN 882
QY 885 NGLSLTGENADTKGLNLTSESATFGKTRDTLNTNGTNNGTAEINTQGVVKLGNTV 944
Db 883 SGNLTAGNVINAGNLTVESNANFKAITFTFNVGGLFDKNGNSISIAKGGARFKDID 942
QY 945 NGDGLNITHAKRNQRTISGDIINKKGSINLTDSDNNDAEIOIGNISOKEGNLTISDD 1004
Db 943 NSKNLSITNSSSTYTLISGNTNKGDLNITNEGSDTEMOIGGDSVSKEGNLTISDD 1002
QY 1005 INTTKQITIKKIDGDESSDTSANLTIKTKELKLTEDLSISGPNKAEITAKGDRLT 1064
Db 1003 INTTKQITIKAGVDGSDSDATNNANLTIKTKELKLTQDLNISGPNKAEITAKGSDLT 1062

QY 1065 IGNSNDGSGAEAKVTYFNNVSKISADGHNVTLSNKVKTSSSNGGRESNDNDGLTI 1124
Db 1063 IGNTNSAD-GYNKAKVTENQVRDSKISADGHKVTLSHKSVEITSGNNNTEDSDNNAGLTI 1121
QY 1125 TAKNVEVNDTSLTKVNITA-SEKVTYTAGSTINATNGKASITTKT----- 1170
Db 1122 DAKNVTYNNNITSHKAVISATSATSGEITTKGTGTINATNGVEITAGTOSILGGTSSSGS 1181
QY 1171 -----GDISGTISGNT 1181
Db 1182 VTLTATEGALAVNSIGNVTVTANSALTLTLAGSTIKGTESVTTSSGSDIGGTISG 1241
QY 1182 VSVSATVDLTTKSGKIEAKGEANVTSATGTIGGTIGTNTVNTANAGDLTVNGAEIN 1241
Db 1242 VEVKATESLTTQSNKIKATTEANVTSATGTIGGTIGTNTVNTANAGDLTVNGAEIN 1301
QY 1242 ATEGATLTATNTLTTEAGSSITTKGOVDLLAQNAGSITAGSINAANVTTLNTTGLTTVA 1301
Db 1302 ATEGATLTSSGKLTTEASSHITSAGGOVNLQAODGSVAGSINAANVTTLNTTGLTTVK 1361
QY 1302 GSDIKATSLVINAADKLNAGDSDTEVNANVNASGSGVTAATSSSVNTIGDNTVN 1361
Db 1362 GSNINATSLVINAADKLNAGDSDTEVNANVNASGSGVTAATSSSVNTIGDNTVN 1421
QY 1362 GLNITSKGRNTVRLGRKEIEVKYIOPGVASVEEVIEAKRVLEKVKDLSDDEERETLAKG 1421
Db 1422 GLNLSKGNLTVLLKGVKIDVYIOPGVASVDEVEIEAKRVLEKVKDLSDDEERETLAKG 1481
QY 1422 VSARVPEVNNTVTNTONEFTTRPSSOVIISGKACFSGNGCARVCTNVADGO 1476
Db 1482 VSAREFIEPNTTIVDTQNEFATRLPSRIVISEGRACFSNGDGTVCVNIADNGR 1536

RESULT 13
US-08-530-198-2
; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: protein

Query Match 67.1%; Score 4967; DB 2; Length 1536;
Best Local Similarity 67.3%; Pred. No. 9.3e-275;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

QY 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRLHALKPLSALLSLGVT 60
DB 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRLHALKPLSALLSLGVT 60

QY 61 SIPOSVLASGLQGMVYVHGTATQVQDGNKTIIRNSVDAILNWKQFNIDQNEWQVFLQENN 120
DB 61 SIPOSVLASGLQGMVYVHGTATQVQDGNKTIIRNSVDAILNWKQFNIDQNEWQVFLQENN 120

QY 121 NSAVFNRTVSNQISQLGILDSNGQVFLINPNGITIGDKDAIINTNGFTASTLDSINENIK 180
DB 121 NSAVFNRTVSNQISQLGILDSNGQVFLINPNGITIGDKDAIINTNGFTASTLDSINENIK 180

QY 181 ARNFTPEQTKDKALAEIVNHGLTVGKGSVNLIGKVKNEGVISVNGSGISLLAQKIT 240
DB 181 ARNFTPEQTKDKALAEIVNHGLTVGKGSVNLIGKVKNEGVISVNGSGISLLAQKIT 240

QY 241 ISDIINPTIYTAAPENEAVALNGLDIFAKGGINVRAAIIIRNOGKLSADSVSKDKSGNIV 300
DB 241 ISDIINPTIYTAAPENEAVALNGLDIFAKGGINVRAAIIIRNOGKLSADSVSKDKSGNIV 300

QY 301 LSAKEGEAEITGGVISAQNOQAKGKLMITGDKVTLKTGAVIDLSKEGGETVILGGDERGE 360
DB 301 LSAKEGEAEITGGVISAQNOQAKGKLMITGDKVTLKTGAVIDLSKEGGETVILGGDERGE 360

QY 361 GKNGIOLAKKTSLEKSTINVSKEKGGFAIYVGDIALDGNINAGSGDIAKTGGFVET 420
DB 361 GKNGIOLAKKTSLEKSTINVSKEKGGFAIYVGDIALDGNINAGSGDIAKTGGFVET 420

QY 421 SGHDLFIKONAIVDAKEWLDFDNVSNIAEATGRSNTSEDDY-TCGNSASTPKRKE- 478
DB 421 SGHDLFIKONAIVDAKEWLDFDNVSNIAEATGRSNTSEDDY-TCGNSASTPKRKE- 478

QY 480 KTLTNTTISNLYKNAWTNITASRLTVNSSINIGSHLTHSKGQGGVQIDGDT 539
DB 479 KTLTNTTISNLYKNAWTNITASRLTVNSSINIGSHLTHSKGQGGVQIDGDT 539

QY 540 ----SKGGLNLTYSGGWDVHKNITLD-QGFLNITA-ASVAFEGGNKARDAANAIVAQ 593
DB 538 TGDTRGANLTYSGGWDVHKNITLD-QGFLNITA-ASVAFEGGNKARDAANAIVAQ 593

QY 594 GTVTITGECKDPRANVSLNGTGKGLNLTSSVNN---LTHNLSGTINISGNTTQTRK 650
DB 591 GTIT-SGNOKGFRFNVSNGTSGGLQFTTTRNKYATNKFEGTLNLSGKVNISMVLPK 649

QY 651 NYS-YWQTSQSHSHWVNSALNLETGANFTF-IKYISSNSKGLTTOYRSSAGVNFNGVNGM 708
DB 650 NESGYDKFKRTYNNLTSLNVSSEGEFNLTDRCGDSAGTLTOPYNLNGISF---NKDT 706

QY 709 SPNLKEGAKVNFPLKPNENMTSPKLP-IRFLANITATGGSVFFDIYANHS---RCAE 764
DB 707 TFNVERNARVNFDPKAPGINKYSSLASVNGNISVSGGSGVDFTLASSNVQTPGVV 766

QY 765 LKMSINISNGANFTLNSHVHVGDDAFKINKDLTINATNSNFSLPOTKDDYDGYARNAIN 824
DB 767 INSKYFNVSSTGSLRFXSTGSKTGFSTKDLTLNATGNNITLQVET--DGMIGKIV 824

QY 825 STYNISILGNNVTGGONSSSITGNTTIEKAANVTLEANNANQOQIRDRVILKGLSLV 884
DB 825 AKKNITFEGGNTFGSRKAVTEIEGNTVINNANVTLIGSDFDNHQ--KPLTIKKOVIIN 882

QY 885 NCSLSLTGENADIKGNLTISATFCKTRDTLNTGNFTNGTAENITQGVVVKGNVT 944
DB 883 SGNLTAGNIVNIAGNLTISATFCKTRDTLNTGNFTNGTAENITQGVVVKGNVT 942

QY 945 NQDGLNTTHAKNRQSIIGDIIINKKGSNLTIDSDNDAEIQGNTSOKEGNLTSSDK 1004
DB 943 NKNLSITTTSSSTYRTIISGNITNKNGLNITNEGSDTEMOIGGVDSOKEGNLTSSDK 1002

QY 1005 INITKOITTKKIDGDBSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKORDLT 1064
DB 1003 INITKOITTKKIDGDBSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKORDLT 1062

QY 1065 IGNSNDGNSGAEAKVTFFNNVKDSKISADGHVNTLNSKVKTSSNGGREGSNDTGLTI 1124
DB 1063 IGNTNSAD-GTNNAKVTFNQVRKDSKISADGHVNTLNSKVKTSSNGGREGSNDTGLTI 1121

QY 1125 TAKNVEVNDKITSKLTVNITA-SEKVTITAGSTINATNGKASITTKT----- 1170
DB 1122 DAKNVTNNNITSHKAVSISATSGEITTKGTITNATNGVEITAQTSILGGIESSGS 1181

QY 1171 -----GDISGTISGNT 1181
DB 1182 VTLTATEGALAVNSISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDITGSGT 1241

QY 1182 VSVSATVDLTTKSGKIEAKGEANVTSATGTIGTIGTISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDITGSGT 1241
DB 1242 VEVKATESLTTQSNKIKATTEGEANVTSATGTIGTIGTISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDITGSGT 1301

QY 1242 ATEGATATATGNTLTTEAGSSITSTKQVLLAQNGSIAGSINAANVTLTGTLTVA 1301
DB 1302 ATEGATATATGNTLTTEAGSSITSTKQVLLAQNGSIAGSINAANVTLTGTLTVA 1301

QY 1302 GSDIKATSGTLVINAKDAKLANGDASGDSSTEVNNAVNASGSGSVTAATSSVNTGOLNTVN 1361
DB 1362 GSNINATSGTLVINAKDAKLANGDASGDSSTEVNNAVNASGSGSVTAATSSVNTGOLNTVN 1361

QY 1362 GLNITSKNGINTVLLKGVKIDVKYIQGVASVBEVIEAKRVLEKVKDLSDEERETLAKLG 1421
DB 1422 GLNITSKNGINTVLLKGVKIDVKYIQGVASVBEVIEAKRVLEKVKDLSDEERETLAKLG 1481

QY 1422 VSAREVEPNNTITVNTQNEFTTRPSQVILSEKACFSSGNGARVCTNVADQ 1476
DB 1482 VSAREVEPNNTITVNTQNEFTTRPSQVILSEKACFSSGNGARVCTNVADQ 1536

RESULT 14
US-08-469-880-2
Sequence 2, Application US/08459880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza ;
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832

```

; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-2

Query Match 67.1%; Score 4967; DB 2; Length 1536;
Best Local Similarity 67.3%; Pred. No. 9.3e-275;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKSGSEKPKRMKVRHLAKPLSAMLLSLGYT 60
Db 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKSGSEKPKRMKVRHLAKPLSAMLLSLGYT 60
Qy 61 SIPOSVLASLQGMVHVGTATQVGDGKNTIIRNSVDAILNKKQFNIDQNMVQFLOENN 120
Db 61 SIPOSVLASLQGMVHVGTATQVGDGKNTIIRNSVDAILNKKQFNIDQNMVQFLOENN 120
Qy 121 NSAVFNRTSNOISOLKGLDNGQVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
Db 121 NSAVFNRTSNOISOLKGLDNGQVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
Qy 181 ARNFTFQTKDALKAEIVNHLITVKGDSVNLIGGKVKNEGVI SVNGGSI SLLAGOKIT 240
Db 181 ARNFTFQTKDALKAEIVNHLITVKGDSVNLIGGKVKNEGVI SVNGGSI SLLAGOKIT 240
Qy 241 ISDIINPTIYSAIPNEAVNLGDI FAKGGINVRAATIRNOCKLSADSVSKDKSGNIV 300
Db 241 ISDIINPTIYSAIPNEAVNLGDI FAKGGINVRAATIRNOCKLSADSVSKDKSGNIV 300
Qy 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLKGVAVIDLKSGEGGETYLGGERGE 360
Db 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLKGVAVIDLKSGEGGETYLGGERGE 360
Qy 361 KGNGIQLAKKTSLEKGGSTINVSKEGGAIVWGDIADLDGNIQAQSGDIATGGFVET 420
Db 361 KGNGIQLAKKTSLEKGGSTINVSKEGGAIVWGDIADLDGNIQAQSGDIATGGFVET 420
Qy 421 SGHDLFIKDNAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDFPTGTG-EASDPKKWSEL 479
Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDFPTGTG-EASDPKKWSEL 479
Qy 480 KTTLTNTTISNLYKNWMTMNITASKRTIVNSSINIGSNLSHLILHSKQGGGVQIDGIT 539
Db 479 KTTLTNTTLESILKGTGVNITANORIVYVNSSINL-SNGSLTLWSEGRSGGVVINDIT 537
Qy 540 -----SKGNGLTIYSGGWVDVHKNIITLD-QGFLNITA-ASVAFEGGNKARDAAAKIVAQ 593
Db 538 TGDDTGRGANLTIYSGGWVDVHKNIISLGAQGNINITAKODIAFEKSGNV-----ITGQ 590
Qy 594 GTVTLTGEGKDFRANVSLNGTGKGLNISSVNN--LTHNLSTGINISGNITINOTRK 650
Db 591 GHIIT-SGNOKGFRFNNSVSLNGTSGLOFTTKTKYATNKEFTNLISGKVNISMVLPK 649
Qy 651 NTS-TWQTSHDHWNVSALNLETGANFYF-IKYISSNKGKLTQYRSAGVNGVNCVNM 708
Db 650 NESGVDKFKGRTYWNLSLUNVSESEFNLTIDSRGSDSAGILTQPYNLNGISF---NKDT 706
Qy 709 SFNLKEGAKVNFKLKPNENMTSKPLP-IRFLANITATGGGVSFFDIYANHSG---RGA 764
Db 707 TENVERNARVNDIKAPIGINKYISLNVAFNCGNISVSGGSDVFTLLSSNNVOTPGV 766

```

```

Qy 765 LKMESEINISGANFTLNSHVRGDDAFKINKDLTINATNSNPSLQTKDDFDYDGVARNAIN 824
Db 767 INSKYFNVSTGSSLRFTKSTGTFSEKDLTLNATGNNITLLOVECT--DGMIGKIV 824
Qy 825 STYNIISILGNNVTLGGQSSSIITGNITIEKAANVTLEANNAPNOQNIIRDRVILKGLSLLV 884
Db 825 AKKNITFEKGNITFGSRKAVTEIEGNTVINNNANVTILGSDFDNHQ--KPTIKKDVIIIN 882
Qy 885 NGSLSLTGENADIKGNLTISESATFKGKTRDTLNTITGNTNNGTAEINITGVVVKLGNV 944
Db 883 SGNLTAGGNIVNIAGNLTVESNANFKAITNFTFNVGGLFDNKGNSNISIAKGARFKDID 942
Qy 945 NDGDLNITTHAKRNORSIIIGDIIINKGSLNITDNNDAEIIQIGNISQKGNLTISDDK 1004
Db 943 NSKNLSITNSSSTYRTIISGNITNKGDLNITNBSGSDTEMOIGGDVSOKEGNLTISDDK 1002
Qy 1005 INITKOITIKKIDGEDSSDATSNANLTIKTKELKLTEDLSISGFNKAETITAKDGRDLT 1064
Db 1003 INITKOITIKAGVDGSDSDATNANLTIKTKELKLTQDLNLSGFNKAETITAKDGSDLT 1062
Qy 1065 IGNSNDGNSGAPAKVTFFNNVVKDSKISADGHNVTLNSKVKTSSSNGGREGSDNDTGLTI 1124
Db 1063 IGNTNSAD-GTNNAKVTFNQVKDSKISADGHKVTYLLSHKVEISGSNNNTEDSSDNAGLTI 1121
Qy 1125 TAKNVEVNKDTISLKVNTITA-SEKVTITAGSTINATNGKASITTKT----- 1170
Db 1122 DAKNVTNNNITSHRAVSIATSGETITKTGTITNATNGVEITAGTQSGILGGIESSGS 1181
Qy 1171 -----GDISGTISGNT 1181
Db 1182 VTLTATEGALAVNSISGNTVTVTVANSALTLTLAGSTIKCTESVTTSSOSGDTGGTISGT 1241
Qy 1182 VVSATVDLTITKSGKIEAKSGEANYTSATGTTIGTISGNTVNTANAGDLFVNGCAEIN 1241
Db 1242 VEKATESITTSOSNKIKATGEANYTSATGTTIGTISGNTVNTANAGDLFVNGCAEIN 1301
Qy 1242 ATEGAATLTATNTLTTEAGSSITSTKGQVDLLAQNAGSSTAGSINAAVNTLTGTLTVA 1301
Db 1302 ATEGAATLTSSKLLTTEASSHITSAGQVNLQAQSGVAGSINAAVNTLTGTLTVA 1361
Qy 1302 GSDIKATSTGLVINAKDAKINGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1361
Db 1362 GSNINATSTGLVINAKDAKINGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1421
Qy 1362 GLNIISKDGRNTVRLKGETEVKYIOPGVASVEEVIEAKRVLEKVKDLSDEREETLAKLG 1421
Db 1422 GLNIISKNGINTVLLKGVKIDVYIOPGVASVDEVEIAKRVLEKVKDLSDEREETLAKLG 1481
Qy 1422 VSAVREVEPNNTITVNTONEFTTRPSOVITISGKACFSGSGNGARVCTNVADGQ 1476
Db 1482 VSAVREPNNTITVNTONEFTTRPSOVITISGKACFSGSGNGARVCTNVADGQ 1536

```

RESULT 15

US-08-728-470-2

Sequence 2, Application US/08728470

Patent No. 5928651

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/728,470
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/302,832
 FILING DATE: 16-MAR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-633
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1536 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE:
 US-08-728-470-2

Query Match 67.1%; Score 4967; DB 2: Length 1536;
 Best Local Similarity 67.3%; Pred. No. 9,3e-275;
 Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

QY	1	MNKYIRLFKSRNALVALVAVSELARGCDHSTKSGKSPARKVRHLAKPLSALLSLGVT	60
DB	1	MNKYIRLFKSRNALVALVAVSELARGCDHSTKSGKSPARKVRHLAKPLSALLSLGVT	60
QY	61	SIQSVLASGLQGMVDVHGTATMVDGKTIIRNSVDAILNWKQFNIDQEMVQFQENN	120
DB	61	SIQSVLASGLQGMVDVHGTATMVDGKTIIRNSVDAILNWKQFNIDQEMVQFQENN	120
QY	121	NSAVFNRTVNSQISQKGLDSNGQVFLINPNTIGKDAIINTNGFTASTLDSINIK	180
DB	121	NSAVFNRTVNSQISQKGLDSNGQVFLINPNTIGKDAIINTNGFTASTLDSINIK	180
QY	181	ARNFTFEQTKDALKAEIVNHEGLITVKGDSVNLGGKVKNEGVIISVNGSISLLAQKIT	240
DB	181	ARNFTFEQTKDALKAEIVNHEGLITVKGDSVNLGGKVKNEGVIISVNGSISLLAQKIT	240
QY	241	ISDIINPTITYSIAAPENEAVALNGLDIPAKGGINVRAATIRNOGKLSADSVSKDSGNIV	300
DB	241	ISDIINPTITYSIAAPENEAVALNGLDIPAKGGINVRAATIRNOGKLSADSVSKDSGNIV	300
QY	301	LSAKEGAETGGVISAQNOQAKGKLMITGDKVTLTKGAVIDLSGREGGFTYLGGERGE	360
DB	301	LSAKEGAETGGVISAQNOQAKGKLMITGDKVTLTKGAVIDLSGREGGFTYLGGERGE	360
QY	361	KGKGIQAKKTSLEKSGTINVSGKEGGFAIVGDTALIDGNINAQSGDIAKTGGFVET	420
DB	361	KGKGIQAKKTSLEKSGTINVSGKEGGFAIVGDTALIDGNINAQSGDIAKTGGFVET	420
QY	421	SGHDLFTKDAIVADAKEWLLDFDNVNSINAEPLFNNTNGINDEPPTGTG-EASDPKNSL	479
DB	421	SGHDLFTKDAIVADAKEWLLDFDNVNSINAEPLFNNTNGINDEPPTGTG-EASDPKNSL	479
QY	480	KTTLTNTTISNLYLNKAWTNMTASRKLTVNSSINTGNSHLLHLSKGQGGQVGDGDT	539
DB	479	KTTLTNTTISNLYLNKAWTNMTASRKLTVNSSINTGNSHLLHLSKGQGGQVGDGDT	537
QY	540	-----SKGGLTIYSGGWDVHKNITLID-OGFLNITA-ASVAFEGGNKARDANAKIVAQ	593

DB	538	TGDDTRGANLTIYSGGWDVHKNISLGRAGGNINITAKODIAFEKGSNOV-----ITQ	590
QY	594	GTVTITGECKDFRANVNSLNGTGKGLNITISSVNN---LTHNLSETINISGNITINOTTRK	650
DB	591	GTIT-SGNQKGFRENNVSLNGTSGLOFTTKRTNKYAITNKEFETLNSGKVNISMVLPK	649
QY	651	NTS-YWQTSKSHSHNVNSALNLETGANFTF-IKYISSNSKGLTTQYRSSAGVNFVNGVNM	708
DB	650	NESGYDKFKGRTYWNLTSLNVSESEFNLITDSRGSAGTLTPYNLNGISF---NKDT	706
QY	709	SFNLKEGAKVNFKLKPNENMTSKPLP-IREFLANITATGGGVFFDIYANHSG---RGAE	764
DB	707	TENVERNARVNFIDKAPIGINKYSLNVAEPNGNISVSGGSDVFTLASSNVQTFGVV	766
QY	765	LKMEINISGANTFLNSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAIN	824
DB	767	INSKYFNVTGSSLRFKTSGSTKTFSTIEKDLTLNATGGNITLLQVEGT--DGMIGKIV	824
QY	825	STYNISILGGNVTLGGGSSSITGNITIEKAANVTLEANNAPNOQNIORVTKLSLLV	884
DB	825	AKKNITEGGNITFGSRKAVTEEGNVTIINANNVTLLIGSDFDNHQ--KPLTIKKDVIIN	882
QY	885	NGSLTGENADIKGNLTISESATFKGTRDTLNITGNFTNNGTAEINITQGVVKLCNVT	944
DB	883	SGNLTAGNIVNAGNLTVESNANFKAITNFENVGGLFDNKGNSNISIAKGGAREKID	942
QY	945	NDGLNITTHAKRNQRSIIGDIIINKGSLNITDSNNDABEIQIGGNISQKEGMLTSSDK	1004
DB	943	NSKNLSITTSSTYRTLIISGNITNKNGLNITNESDTEMOIGGVDVSQKEGMLTSSDK	1002
QY	1005	INITKQITIKKIGDGEDSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKGDRILT	1064
DB	1003	INITKQITIKAGVDGSDSDATNANLTIKTKELKLTQDLNLSGFNKAETAKGSDLT	1062
QY	1065	IGNSDNGSCAEAKTYTFENVKDKLSADGHVNTLSKVKYKTSNNGGSRNSDNDGLTI	1124
DB	1063	IGNINAD-GTNAKKVTFNQVKDKLSADGHVNTLSKVKYKTSNNGGSRNSDNDGLTI	1121
QY	1125	TAKNEVNKDIITSLKTVNITA-SEKVTITAGSTINATNGKASITTKT-----	1170
DB	1122	DAKNVTNNITSHKAVSISATSGETITTKTGINATTGNVEITAGTGILGIESSSGS	1181
QY	1171	-----GDISGTISGNT	1181
DB	1182	VLTATGALAVNSISGNTVTVTANSALTLTLAGSTIKGTESVTTSSQSGDIGTISG	1241
QY	1182	VYSATVDTLTGSGKIEAKSGEANTVTSATGTIGTISGNTVNTANAGDLTVNGAEIN	1241
DB	1242	VEKATESLTOSKIKATGEANTVTSATGTIGTISGNTVNTANAGDLTVNGAEIN	1301
QY	1242	ATEGAATLTATGNTLTTEAGSSITTKGOVDLLAONGTAGSINAANTVNTTGLTTVA	1301
DB	1302	ATEGAATLTSSGKLTTEASSHITSAGQVNLASAQSGVAGSINAANTVNTTGLTTVK	1361
QY	1302	GSDIKATSGTLVINAKDAKNGDASGDSDEVNAVNASGGSVTAATSSVNTIGDLTVN	1361
DB	1362	GSNINATSGTLVINAKDAKNGDASGDSDEVNAVNASGGSVTAATSSVNTIGDLTVN	1421
QY	1362	GLNIIKSGDRNIVRLRKEIEVKYIQGVASVEEVIEAKRVLEKVKDLSDERETAKLG	1421
DB	1422	GLNIIKSGDRNIVRLRKEIEVKYIQGVASVEEVIEAKRVLEKVKDLSDERETAKLG	1481
QY	1422	VSARFVPEPNTITVNTQNEFTTPSSQVILISEGKACFSSGNGARCVTNVADGQ	1476
DB	1482	VSARFVPEPNTITVNTQNEFTTPSSQVILISEGKACFSSGNGARCVTNVADGQ	1536

Search completed: March 24, 2003, 15:24:58
 Job time : 28 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run On: March 24, 2003, 15:21:24 ; Search time 18 Seconds
(without alignments)
3403.363 Million cell updates/sec

Title: US-10-092-880-4
Perfect score: 7407
Sequence: 1 MNKIYLFKSKRLNALVAVS.....CFSSGNGARVCTINVADGGP 1477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	555.5	7.5	1577	1 HLYA_PROMI	P16466 proteus mir
2	521	7.0	2003	1 YDPA_ECOLI	P33666 escherichia
3	511	6.9	2249	1 OMPA_RICRI	P15921 rickettsia
4	509	6.9	917	1 HXA3_HAEIN	P45355 haemophilus
5	486.5	6.6	928	1 HXA2_HAEIN	P45354 haemophilus
6	476.5	6.4	1656	1 OMPB_RICJA	O06653 r outer mem
7	475.5	6.4	2021	1 OMPA_RICGN	O02657 rickettsia
8	469.5	6.3	3591	1 FHAB_EORPE	P12255 bordetella
9	466	6.3	905	1 HXAL_HAEIN	P44602 haemophilus
10	441.5	6.0	1325	1 YDEK_ECOLI	P32051 escherichia
11	429.5	5.8	1654	1 OMPB_RICRI	Q3047 r outer mem
12	429.5	5.8	1953	1 BIGA_SALTY	P25927 salmonella
13	428	5.8	1655	1 OMPB_RICCN	Q9Kka3 r outer mem
14	425	5.7	1608	1 HLYA_SERMA	P5320 serratia ma
15	419	5.7	1643	1 OMPB_RICPR	Q53020 r outer mem
16	415	5.6	1645	1 OMPB_RICTY	P96989 r outer mem
17	406	5.5	2660	1 YEEJ_ECO57	Q8x8v7 escherichia
18	402.5	5.4	1286	1 AIDA_ECOLI	O03155 escherichia
19	385.5	5.2	2358	1 YEEJ_ECOLI	P76347 escherichia
20	363.5	4.9	1569	1 YRJA_ECOLI	P52143 escherichia
21	358	4.8	1754	1 OMPB_CHLTR	O84418 chlamydia t
22	357	4.8	1694	1 IGA0_HAEIN	P44969 haemophilus
23	355	4.8	1672	1 OMPB_CHLMU	P38058 clostridium
24	354	4.8	1848	1 CBPA_GLOCL	P45384 haemophilus
25	352	4.7	1702	1 IGA2_HAEIN	O92812 chlamydia p
26	346.5	4.6	1723	1 PMW0_CHLPN	O84419 chlamydia t
27	344	4.6	1770	1 PMPC_CHLFR	P14914 rickettsia
28	334.5	4.5	1500	1 I20K_RICRI	P45828 caulobacter
29	333.5	4.5	1025	1 SLAP_CAUCR	P45508 escherichia
30	324.5	4.4	1250	1 YFAL_ECOLI	Q02470 lactobacilli
31	322	4.3	1902	1 P2P_LACPA	P15293 lactococcus
32	319	4.3	1902	1 P2P_LACLC	P15292 lactococcus
33	318.5	4.3	1902	1 P3P_LACLC	

34	317	4.3	1341	1 VG37_BPT2	P07067 bacterioph
35	315.5	4.3	1310	1 VAC3_HELPY	O48253 helicobacte
36	314	4.2	1253	1 D5PP_HUMAN	O9nz44 homo sapien
37	314	4.2	1902	1 P1P_LACLC	P16271 lactococcus
38	313	4.2	1933	1 SLAP_CAMFE	P35827 campylobact
39	308	4.2	1567	1 ICEN_XANCT	P18127 xanthomonas
40	297	4.0	1849	1 IGA4_HAEIN	P45386 haemophilus
41	296	4.0	1541	1 IGA1_HAEIN	P42782 haemophilus
42	294	4.0	1861	1 APU_THETU	P38536 t amylopull
43	290.5	3.9	1243	1 VG37_BPK3	Q38394 bacterioph
44	288	3.9	1545	1 IGA3_HAEIN	P45385 haemophilus
45	284.5	3.8	1176	1 SLAP_BACSH	P38537 bacillus sp

ALIGNMENTS

RESULT 1
HLVA_PROMI
ID HLVA_PROMI STANDARD; PRT: 1577 AA.
AC P16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN HPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC STRAIN=Isolate 477-12;
RX MEDLINE=90170827; PubMed=2407716;
RA Uphoff T.S., Welch R.A.;
RT Nucleotide sequencing of the proteus mirabilis calcium-independent
RT hemolysin genes (hpmA and hpmB) reveals sequence similarity with the
RT serratia marcescens hemolysin genes (shla and shlb).
RT J. Bacteriol. 172:1206-1216(1990).
RL -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC DEFINED.
CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEMP-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
CC REQUIRES HPMB FUNCTION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA
CC MAY BE RESPONSIBLE FOR PORE FORMATION.
CC -1- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30186; AAA25657.1; ..
CC PIR; A35140; A35140.
CC Hemolysis; Toxin; Outer membrane; Signal.
CC SIGNAL 1 29
CC FT CHAIN 30 1577 HEMOLYSIN.
CC SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

Query Match 7.5%; Score 555.5; DB 1; Length 1577;
Best Local Similarity 20.7%; Pred. No. 1.1e-15;
Matches 361; Conservative 257; Mismatches 606; Indels 523; Gaps 79;
Oy 40 MKVRHLAKPLSAMLISLGVTSIPQSVLASGL-----QGMDVV---HGTATMQV---DGN 88
DB 1 MKSFKFLSPSGRLAASLAIFVSLNAYGNIGVDPAGHQGPDSVAVNGGTQVINIVTPNN 60

```
QY 89 KTIIRNSVDAILNKKQFNI-----DONEMVQFLOENN-----SAFNRVT 129
Db 61 EGISHN-----QYDFNVGPKGAVFNNALEAGOSQAGHLNANSNLNGQAASLIILNEV 114
QY 130 SNOISQLK---ILDSNGQVFLNPNGITIGKDAINTNGFTASTLIDISNENIKARFTF 186
Db 115 SRNPSFLGQOEVFGIAAEVLSNPNGITCDGCGFIN-----SRSSLVVGNPFLF 164
QY 187 EYTKDKALAEIVNHLITVGD-----GSYNLGGKVKNEGVISVNGGSIISLAGQKITI 241
Db 165 ENGQKGYSTLNTNLLSLGKNGLNTTGLDLIAPRDSRG--KITAAEISAFNG--TF 221
QY 242 S---DIINPITYS-----IAAPNEAVNLGDIIPAKGNNINVAATIR 281
Db 222 SQHFDILSSOKPVSALDSYFFGSMQSGRIIRINTAEGSGVKLAGFTADNDLSVKADNIQ 281
QY 282 NQKLSADSYSKDKSGNI-----VLSAKEGEAEITGGVISAQOAKGK 325
Db 282 TDSQVRYDSYDKGSENYQNYRGITVNNSSQTLTKTELKGNITLVASHNQIKASD 341
QY 326 LMITGDKVTLKTCVADLDSKEGGETYLCGDERG--EGKNGIOLAKKTSLEKGSTINYSG 383
Db 342 LM--GDIITLQ--GADLITIDGKLOQKETDIDNRWFYSWKYDVTKEKEQIQQIGSID--A 396
QY 384 KEGGFAIYMGDIALDIGNINA-----QSGDIAKTG----- 415
Db 397 KNNATLTATKGDVTLDAAKINAGNNLAINANKDIIHNGLVEKESSENGKRNHTSRLES 456
QY 416 -----GFTVSGHDLFIKDNAIVDAKEMLLDFDNYSI 447
Db 457 GWSNSHOTETIKASBELTAGDLGDAQGSITAOQAKLANENVLVNAK-----DNINL 510
QY 448 NAEDPLFNNTGIDNDEFTGTGEASDPKKNSELKTLTNTTISNLYLKNWNTNITASRLT 507
Db 511 NVQ--KTNDKTYTDNHVMWGGIGGQCNKNNNQOVSHAT-----QLTADGQL 558
QY 508 VNSSNIGSNHLIILHSKQRG-----GGVQIDGDTITSKGNLTIYSGWVDV---H 556
Db 559 LAADNNVAITGSQV---KGNQAFVKTQTQGDVVIDNALSETISKIDERTGAFNLTSSH 615
QY 557 KNITLDQGLNITAASVAFEGNGKNKARDAANAKIVAQGVITITGEGKDFRANVSLNGTG 616
Db 616 KNET-----NK-QTSTGSELISDAQLTWS-----GNDVNVIG-- 647
QY 617 KGLNIISSVNNLTHNLSTGNI-----SGNITINQ 646
Db 648 ---SLIKSADKLGIHSLGDIINVSAQAQVTKIDDEKTSIAITGHAKEVEDKQYSAGPHITH 704
QY 647 TTRKNYSYQWQTHSHWNVSALNLETGAN--FTF-----IKYISSNS 686
Db 705 TTNKNTS--TETEANSTISGANVLDQANKDVYTFAGSDLKTATAGNASITGDNVAFVSTEN 762
QY 687 KGLT--TOYRSSAGVNF--NGVNGNMSFNLKBGAKVNF---KLKPNENMNTSKPLPIRELA 740
Db 763 KKQTDNTDTTISGFSYTGVD-----KVGSKADQYQKHQHTQTEVTNRKGSQTEVAG 815
QY 741 NITATGGSVFDDIYANH--SGRGAEKMSFINISNGANFTLNSHVGRDDAFKINKDLTIN 799
Db 816 DLTITANKDLLHEGASHHVEGRYOE--SGENIGHLA--VNDSETSKTDSLNVGIDVGN 870
QY 800 ATNSNFSLRQTKDFDYGY-----ARNAINSYNITSIL-----GGNVTL 838
Db 871 LDYSGVT--KPKVKAIEDGVNVTKEPNNVTDTTKKVARTADALANLANLSNLETTPNVGVEGI 929
QY 839 GGQSSSITGNTITIEKAAN---VTLAANNAPQONTRDRVIKLG--SLLVN----- 885
Db 930 KGGGQSQSOTDSQAVSTSINAGKIDIDSNNKLHDQGHYQSTQEGISLTANTHTSEATLD 989
QY 886 -----GSLSLTGEN--ADIKGNLITISESATFFKCKTRDTNLITGNFTNNGT 928
Db 990 KHQTFTHETGGGQIGVSTGTGSDITVAIKGEGQTTDNALMETKAGS-----QFTSNGD 1044
QY 929 AEINITQGVVVLGNTVNDGDLNITTHAKRNQRSIIIGDDIINKKGSINI---TDSNNDAL 985
```

```
Db 1045 ISINVE-----NAHYE-----AQFDAQK--GKTVINAGGDLTLAQATDTHSESQS 1089
QY 986 QICGNISOKEGNLTISD-----KINITKQITIKKIGDGEDSSSDATSNANLTIKT 1036
Db 1090 NVNGSANLKVGTTPESKDYGGGFNAGTTHHSKEQTTAKVGTITGSQIELNAGHNLITQ 1149
QY 1037 RELKLTEDLSISGPNKAEITAKGDRDLTIGNSNDGNSGAE---KTVTFNNVKDKISAD 1093
Db 1150 THLSSEQDIALNATNKVDLQASSEHTEKGNLSG--GVQAGFGKKMT-----DDASSVN 1202
QY 1094 GHNVTLNS-----KVTSSNGGREGSNDNT--GLTITAKNVENKRDITSL----- 1138
Db 1203 G-----LGSQAFAIGKQDEKSVSREGGITNNSGNLTINGSVHLOGAQVNSKDTQLTSQS 1258
QY 1139 -----KTVNITASEKVTTTAGSTINATNGKASITT---KT 1170
Db 1259 DIEITSQSTDYKNNWGTDDIGFNGKKTNTPEKTEEPATSIHIGKLLLVNVEDQOKT 1318
QY 1171 GDISGTISGNTVSVSATVDLTTKSGSKEAKSGEANTV--SATGTIGGTIS----- 1219
Db 1319 SHQATLETGTLTINSKDLTL-----SG--ANVTADSVTGNVGGSLNIASQKESDR 1368
QY 1220 -----GNTVNVNTANAGDL-----TVGNGAEINATEGAATLTATGN 1254
Db 1369 HVTGVNVGVNHTNDPKSSQVNTAKAGGSILLEKTIKDTIDSGIKSSTDALISOKYNSLS 1428
QY 1255 TLTTAEAGSITSTKGQVDDLLAONGSIAGSINAAVNTLNTTGTTLTVAGSDIKATSGFLVI 1314
Db 1429 TIADKTGIS--DETKAKID--QGFGKVGNGIK--NI---VTGAEGHTANADIKVT----- 1474
QY 1315 NAKDAKLNAGSDSTEVNAVNASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTV 1374
Db 1475 -----HVDNDVAVTKT-----SLTNSNDLSLVNGS--TKLTGAEIVSQQGG--V 1515
QY 1375 RLGRKELEVYKIQGVASVEEVTEAKRVLEKVKDLSDEERETLAKLGVA--VRFVEPNNT 1433
Db 1516 DLGGSSVKLENIE-----GHHYEAGADLDLKSSVVD-----LAKQLVGGDISFKSPVKT 1564
QY 1434 -ITVNTQ 1439
Db 1565 NETVNTK 1571
```

RESULT 2

```
YDBA_ECOLI
ID YDBA_ECOLI STANDARD; PRT: 2003 AA
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
```

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasubramaniam S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.,
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Meszer I., Glaser P., Danchin A.;
RT "Multiple *Es. coli* K-12" insertion sequences near the replication terminus in
RT *Escherichia coli* K-12";
RL Biochimie 73:1361-1374(1991).
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYS6 (AC P25928).
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000237; AAC744483.1; ALT-SEQ.
CC EMBL; AE000237; AAC744487.1; ALT-SEQ.
CC EMBL; D90778; BAA15009.1; ALT-SEQ.
CC EMBL; D90778; BAA18880.1; ALT-SEQ.
CC EMBL; D90779; BAA18881.1; ALT-SEQ.
CC EMBL; X62680; -; NOT_ANNOTATED_CDS.
CC EcoGene; EG11307; ydbA.
CC Hypothetical protein; Complete proteome.
CC I -> V (IN REF. 2).
CC FLICCT 489 489 I -> V (IN REF. 2).
CC FLICCT 495 495 I -> V (IN REF. 2).
CC SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;
CC -----
CC Query Match 7.0%; Score 521; DB 1; Length 2003;
CC Best Local Similarity 21.7%; Pred. No. 3.6e-14;
CC Matches 388; Conservative 229; Mismatches 625; Indels 542; Gaps 90;
CC -----
CC 89 KTI-IRSVDAIINWKFIDONE--MVQFLQENNNNSAVENR---VTSNOISOLKGLTDS 142
CC 153 KTLIRSV-----FTYENADGTISLQDSNGRKATINLMQIDEANNTVALEGVSD 204
CC 143 NGQVFLNPNG--ITIGKDAIINTNGTASTLDLISNENIKARNFTFQTKDKALAEIVNH 200
CC 205 GATKQVYHNHGLVITGDATVNNNGKT--TVD----- 235
CC 201 GLITVQKDGSG---VNLIGGKVKNGVTSVNGSGISL-LAGOKTITISDIINPTITYSTAAP 256
CC 236 -----GKDSSTGEINGNNGKVIQDGLDVGSGHGHIDITGSATVDNKGMTVT-----DP 286
CC 257 ENEAVNL-GD-----IFAKGG-----NINRAATIRNOGKLSA---DSVKDKSGNI 299
CC 287 ESMGIQIDGDKRAIVNNEGESTITNGGTQINGDDATANNNGKTTPDGKDSGTGTEINGNN 346
CC 300 VLSAEGAEIGGVISAQNOAKGKLMITGDKVTLTKTGAVIDLSGREGGTYLGGDE-- 357
CC 347 KVIQDGLDVGSG-----GGHGHIDITGSATVDNKGMTVTDPESIGIQVDGQAV 397
CC 358 -RGEKNGIOLAKTISLEKGTIN-----VSGKEGKGFAIVWGDIALIDGNI 403
CC 398 VNNGESAI-----TNGGTGTQINGDDATANNNGKTTPDGKDSGTGTEIAGNNGKVIQ--- 449
CC 404 NAQSGSDIAKGGFVETSGHDLFIK-DNAIVDAKEWILDFD-----NVSINAEPLFNNTG 458
CC 450 -----DGLDVGSG-----GHGHIDITGSATVDNKGMTVTDPESIGIQDQDAIVNNEG 500
CC 459 INDEFPTGTG-----EASDPKKNSELKTTL-----TNTTISNLYKNATWMTNITASRKLTVNS 510

Db 501 ESTITNGGTGTQINGNDATANNCKTTVDGKDSGTGTGIAG---NIGIVNLDGS--LVTGT 555
Qy 511 SI-----NIGSNHLLHLSK-----QRGG-CVQIDGDTTSKGGNLTIVSGWVDVHK--- 557
Db 556 GAHGVENIGDNG--TVNKKGDIVVSDTSGIGVLNGE-----CATVSTNGTDVNVSNNEAT 607
Qy 558 --NITLDGFLNIT-----AASVAPEGCGNKKARDA--NAKIVAO-----GT 595
Db 608 GFSITTSKGVSLAGSMQGVDFSTCVDLNGNNSVTLAAKDLKVVQKATGINVSGDANT 667
Qy 596 VTIGE---GKDFRANVVS--LNGTGKGLNISSVNNLT-----HNL- 632
Db 668 VNITGNVLVDKDTADNAAEYFFDPVGINVIGSDNNVTLTGKLVSDSEVTSRQSNLF 727
Qy 633 -----SGTINISGNTI-----NQTRKNSTYQWTS-----HDS 661
Db 728 DGSAEKTSGLVVGDTNVMNMGLELIGEKNALADGQVTSRLRTGYSTVIVVSGESS 787
Qy 662 HW-----NVSALETTGANFTFIKYISSNSKGLTQVRSAGV 699
Db 788 VYLNQDGTITISGEPFLGFAGVIRVODKALLEIGSGATLTMODIDSEFHHGTRTVEIQNLGF 847
Qy 700 NENGYNMNSFNLKEGAKVNEFKLPENNNMTSKPLPIRFLANI--TATGGSVFFDIYAN 757
Db 848 AF--VTGENTTIGNSGT-----ISLLQNGKDPAPSPVILLATNGGSGATNAGTITGKVTEQ 900
Qy 758 HS-----GRGA--ELKMSSEIN 771
Db 901 HSFVFNKYSTGTSNFIENNDVSSITGLVAQSNSTIINTDSTGIDLYGRGVCLMAIDST 960
Qy 772 ISGANFTLNS--HVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINTSYNLS 830
Db 961 AENQCKITLDSMWVDANDTTAM--RDIASNSAIDFGVGVGTDSISGAKKNAT-----AIN 1015
Qy 831 ILGGNVT-----LGGQNSSSI--TGNITTEK-----AANVTL--EANNAPN 868
Db 1016 QLGGVITTYNAGAGMAAYCASNTVINOGTINLEKNGNYDLSAANTLVGMAYVEHGTAIN 1075
Qy 869 QQ-----NIR-----DRVTKGLSVLVNG-----SLSTENADI 897
Db 1076 DQTVGININVTGQAFYNDGTGTIVNYGPTCTFGVCQSQNEYNNTDDFTSLYITGDTTIT 1135
Qy 898 KG--NLITISATFPGKTRDTLNTGNFTNNGTA---EINITQGVVVKLVNVDGDLNIT 952
Db 1136 RSGETVTLNKSAAVTDK-----LAGNVVNSGTLSDQITVSSGLE---NTSGGIIN-- 1184
Qy 953 THAKRNORSITIGDIIKKKGLNITDSNNDAAETIQIGNISQEGNLTISDKNITKQIT 1012
Db 1185 -----NLVKLDKGAIVKNAGVM---TNN---VDVSGGILNNAGEM-----TAQIT 1223
Qy 1013 IKKGIDGEDSSSDATSN---ANLTKTKELKLTEDLSISG--FNK-----AEIT-- 1056
Db 1224 MNAGDSSLVNTGNTINKIVQNAVFNNSGVSVTGRMMSAGGVFNQNTDGAIRGAALTCT 1283
Qy 1057 --AKGDRDLTIGNSDNGSGAEAK---TVTFENNVKDSKISADGHVNTLNS----- 1101
Db 1284 AVANNECTWNLGSSSENGNTGMLNVNNSAFNNRGEFILDNDKNVAHQSGTLYNTGHM 1343
Qy 1102 KVTSSNG-----GRESNDNDTGLTITAK-----NVEVANKDIT 1136
Db 1344 NISNSHNGAVNMMWNGNRFINDGTIDVSAKSLVSVANNAGDONAFWQNDNGVINFDHD 1403
Qy 1137 SLKTVNITASEKVTYTTAG--STINATNGKASITTKTGTDI---SGTIS-----GNT 1181
Db 1404 SASAKVYTHSNFTAQNDGIMNISGTGAVAMEGDKNAQLVNNNGTINLTAGTTDTGTGMQ 1463
Qy 1182 VSVSATVDLTITKSGSKTEAKSGEANTVSATGTIC-----GTI-----SGNTV 1223
Db 1464 LDANATADAVIENNGTINIFANDSFAPSVLGTGVHVNNNGTVVIADGVTSGLIKQDGI 1523
Qy 1224 NVTANAGDLVVGAEINATGEGATLTATGNTLT-----EAGSSITSTKGQVLDLAQNG 1278

Db 1524 NVEGMNCG--NGNSSEVHYGDYTLDPVPKNTVSVTSGSDEAGSGMNNLNGYVVGTVNWG 1581

Qy 1279 SIAGSINAANVTNL-----TTGTLTT-----VAGS---DIKATSGTLVINAKDAK 1320

Db 1582 S-AGKLVNNAANGVBEINTGFTAGTADTTVSFDNVVSGNLTDADAITSVV--WTAK 1638

Qy 1321 LMGDASGDSDEVNAVNASGSGSVTAATSSVNITG----DLNTVGNLNIISKDGRNTVRLR 1377

Db 1639 GSTDASGNDV---VTMSKNAYTDVATDASVNDIAKALDAGYTNNELFTSLNVGTTAEIN 1694

Qy 1378 GKIEVKYIQPGVASVEEIVAEKRVLEKVDLSDEBERETLAKLGVSAREVEPENNITVN 1437

Db 1695 SALKQVS---GSOATTVPREARVLSNRFSLMADAPKVCNGNGLAFNVVAKGDPRAELGN 1750

Qy 1438 TQNEFTT-----RPSQVSIISGKACFSSGNGARVCTNVADG 1475

Db 1751 TEYDMLALRKTDLSQSOTMSLEYGIARLDGGAQ---RAGDNG 1791

RESULT 3

OMPA_RICRI

ID OMPA_RICRI STANDARD; PRT: 2249 AA.

AC P15921;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmp A).

DE OMPA.

OS Rickettsia rickettsii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=783;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R.

RX MEDLINE=90354033; PubMed=2117568;

RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;

RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";

RL Infect. Immun. 58:2760-2769(1990).

CC 1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.

CC 1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A SLAYER WITH HEXAGONAL SYMMETRY.

CC 1- PTM: GLYCOSYLATED (PROBABLE).

CC 1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: M31227; AAA26380.1; .

DR PIR: A41477; A41477.

DR InterPro: IPR003858; rOmpA_rOmpB.

DR Pfam: PF02708; rOmpA_rOmpB; 1.

DR Antigen; Repeat; Signal; Cell wall; s-layer; Glycoprotein.

FT SIGNAL 1 28

FT CHAIN 29 2249

FT POTENTIAL.

FT OUTER MEMBRANE PROTEIN A.

FT DOMAIN 212 1180

FT REPEAT 212 286

FT REPEAT 287 358

FT REPEAT 359 430

FT REPEAT 431 505

FT REPEAT 506 577

FT REPEAT 578 652

FT REPEAT 653 724

FT REPEAT 725 799

FT REPEAT 800 874

FT REPEAT 875 949

FT REPEAT 13 X APPROXIMATE TANDEM REPEATS.

FT A (TYPE I).

FT B (TYPE II).

FT C (TYPE II).

FT D (TYPE II).

FT E (TYPE II).

FT F (TYPE I).

FT G (TYPE I).

FT H (TYPE I).

FT I (TYPE I).

FT J (TYPE I).

FT REPEAT 950 1021 K (TYPE II).

FT REPEAT 1022 1093 L (TYPE II).

FT REPEAT 1094 1165 M (TYPE II).

FT REPEAT 1166 1180 TYPE I (INCOMPLETE).

SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match 6.9%; Score 511; DB 1; Length 2249;

Best Local Similarity 22.1%; Pred. No. 1.1e-13;

Matches 335; Conservative 197; Mismatches 593; Indels 388; Gaps 68;

Qy 119 NNSAVNRVTSNOISQKGLDSNGQVFLINPNG-----ITIGKDAIINTNGFTAST 171

Db 47 NNAAFSNVGNNNNNEITAAGVANG-----TPAGQPNNWAFYGGDYV-----T 93

Qy 172 LDISNENIKARNFTFEQTKDALKAEIVNHLITVVGKDSVNLGGKVKNEGIVSVNGSI 231

Db 94 ADAADRIIKAINVA-----GTPVGLNITQNTVGSIIITKG---NLLPV 134

Qy 232 SLLAGQKITISDIINPTITVYSIAAPENAVNLGDIIFAKGGINVRAATIRNOGKLSADSV 291

Db 135 TLNAGKSLTLNGNNAVAANHGFDPADNVTGLGNIALGGAN---AALI-----I 180

Qy 292 SKDKSNIVLSAKEGEAEITGVISAQNOQAKGKLMITGDKVTLKTGAVIDLSKGEGET 351

Db 181 QSAAPSKITLA---GNIDGGIITVKTDAIINGTIGNATATVNVGA-----GTA 228

Qy 352 YLGG-----DERGEGKNGIOLAKKTSLEKSGTINVSKEKGFPAIVGDIADIDGNIN 404

Db 229 TLGGAVIKATTIKLTNAASVLTNANAVLTGAIDNTGDNVGNLNGALSQVTDIG 288

Qy 405 AGSGDIATGGFVETSGHDLFIKDAIVDAKEWLLDFDNVSINAEPLFNNTGINDPEP 464

Db 289 NTNSLATISVGAGTATLG-----GAVIKATTIKLTDAASAVKFTNPVVVTAIDN--- 338

Qy 465 TGTGEASDPKNSKELKTLTNTTISNYLKNA---WTMTITASKRLTVN-----SSINTG 515

Db 339 ---TGNA-----NGVFTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTG 390

Qy 516 SNSHILHSGKRGQGGVQIDGDIITSGK-----GNLTIYSGGWVDPVHKNTILDQCF 565

Db 391 DNASAVTFT-----NPVVVTVTAIDNTGNANNGIVFTGNTSV-TGDIGNATATVNVGA 444

Qy 566 LNTAASVAFEGGNKARDAAAKIVAOGTVITGEGKDFRANNSVLTGKG-LNIIS 624

Db 445 GTATLGGAVIKATTIKLTNAASVLTNANAVLTG-----AIDNTGDNVGNLNGA 498

Qy 625 VNNLTHLSGTINIS-----GNITINQTTKNTSYWQTSKSHSHSNVYALN--LE 671

Db 499 LSQVTGNTGNTSLATISVGAGTATLGGAVIKATTIKLT-----DAASVKTNPVVV 551

Qy 672 TGANFTFKYISSNSKGLTTOYRSSAGVNFNG---VNGNMSPNLKGAQVNFKLKPENNM 728

Db 552 TGA-----IDNTGNANNGIVT-----FTGNSTVTDIG----- 579

Qy 729 NTSKPLPIRFLANI-----TATGGSVFEDLIYANHSRGAEKLMSEIN- ISGANFTLNS 782

Db 580 NTNS-----LATISVGAGTATLGGAVIKATTIKLTNAASVLTNANAVLTGA---IDN 630

Qy 783 HVRGDDAFKINKDLTINATNSFSLROTDDFYDGYARNAINSTVNIISILGNVTLGG-- 840

Db 631 TTGGDN-----VGVNLNGLALSQVTDIGI-----GNTNSLATISVGAGTATLGGAV 675

Qy 841 -QNSSSITGNITTEKAANVTLEANNAPNOONIRDRVILKGLSLVNGSLSTGENADIKG 899

Db 676 IKATTTKITNAVSAVKFTN-----PVVVTGAIDSTG-NAN-NG 711

Qy 900 NLTISESATEFGKTRDTLNTITGFTN-NGTAEINITOGVVKLGNVTNDGDLNITHAKRN 958

Db 712 IVTFTGNST-----VTGDIIGNATATVNVGAGTATLGGAV---IKATTTKLTN 757

Qy 959 QRSIIIGDIIKKKGLS-----NITDSNDAEIOIGGNISOKEGNLTISSKINIKTQITI 1013

Db 758 AASVLA--TLTNANAVLTGAIDNTGDNVGNLNGALSQVTDIG-----NTNSLATI 809


```

Db 559 -----LDLDRYQLS-----ELEQRR-RWRYRDLDLDMNKAYLYR----- 594
QY 742 ITATGGGVDFDIYA-NHSGR---GAELKM--SEINISNGANFTLNHVRGDDAFKINK 794
Db 595 -----FDLFAKNSSRTKTEINISNKLNGFVHLLAENIKLDS---KV 641
QY 795 DLTINATNSFSLRQTKDDFDYGVARNAINSTYNISLGG---NVTLGGQNSSSSTGNT 851
Db 642 DITFDKNSQDTLAQTNRLGMNGKV-SMINS--HIKIVGDEKIGISPTGYATMFLIGEL 698
QY 852 TIEKAA-----NVTLEANNAPNOINRDRVILKIG-----SLLVNGSLSLT 891
Db 699 IGEKSSIFVKSHOGYTFKTDGDKTKIAGKN--SKEDLKITAINTGGRAAEVLLINGALSA 756
QY 892 GENADIKG-NLTISGATFGKTRDTLNITNGTNGTAIEINITQGVVVLGNVNDGDUN 950
Db 757 DNDANIANMAFTICDSANTY-TTIENADITALAPNGGTAYLSKDVIEIVKPSNFTFFE 815
QY 951 IITHAKRNQRSIGGDI-INKKGSLNITDSNNDAEIQIGGNSIQKQEGNLTISDDKINIK 1009
Db 816 LPREKLNQTKINGASTKLSEKGFARLYDKINGVR---ASNLSAQLNVTDASEKIINTK 872
QY 1010 QITIKKRGIDGEDSSSDATSA 1030
Db 873 LVS---SLDVEKLVSVAUCDA 890

RESULT 5
HX42_HAEIN
ID HX42_HAEIN STANDARD; PRT; 928 AA.
AC P45354;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
DE protein A).
GN HXA4.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-35.
RC STRAIN-DL42 / Serotype B;
RX MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.;
RT "The 100 kDa haem:haemopexin-binding protein of Haemophilus
RT influenzae: structure and localization.";
RL Mol. Microbiol. 13:863-873(1994).
RN [2]
RP SEQUENCE OF 1-30 FROM N.A.
RC STRAIN-DL42 / Serotype B;
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yorgev R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
RT heme:hemopexin by Haemophilus influenzae type b.";
RL J. Bacteriol. 177:2644-2653(1995).
CC -!- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U08348; AAA74138.1;
CC Transport; Signal; Repeat.
KW SIGNAL
FT CHAIN 1 21
FT CHAIN 22 928 HEME/HEMOPEXIN-BINDING PROTEIN.

```

```

FT DOMAIN 101 679 6 X 6 AA APPROXIMATE REPEATS.
FT REPEAT 101 106 1-1.
FT REPEAT 205 210 1-2.
FT REPEAT 279 284 1-3.
FT REPEAT 410 415 1-4.
FT REPEAT 635 640 1-5.
FT REPEAT 674 679 1-6.
FT DOMAIN 149 172 4 X 6 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 149 154 2-1.
FT REPEAT 155 160 2-2.
FT REPEAT 161 166 2-3.
FT REPEAT 167 172 2-4.
SQ SEQUENCE 928 AA; 101228 MW; 67D45466A4B92390 CRC64;

Query Match 6.68; Score 486.5; DB 1; Length 928;
Best Local Similarity 23.88; Pred. No. 3.7e-13;
Matches 249; Conservative 166; Mismatches 363; Indels 267; Gaps 49;

QY 77 VHGTATMQ--VDGNKTIIRNSVDAIINWKQFNIDQNMVQFLOENNSAVENRVYSNQIS 134
Db 33 VFGTVTIEKTTADKMTIKQGSDDKAQIDWKSEFICQKKEVFEQNEHAYVNRVIGNAS 92
QY 135 QLKGLDSNGOVFLINPGITIGKDAIINTNGFTASTLDIS--NENIKARNTFEQTKDK 192
Db 93 QIQGKLTANGKVLNPGVITITQGAENVAGLIATTKLERISENSNSYQFT--RRTKDR 151
QY 193 ALAEIVNHGLITTVGKDGSVNLIGKVKNEGVS-----VNGGSISLLAOKKIT 240
Db 152 ---QVLKEGLVL--KDGQV-VKEGVINEGNIITAOQFVNLGDEVINKGINYEKNSTIN 205
QY 241 ISDIINPTIYSTAAPENE-AVNLGDIFAKGGINVRAATIRNQGKLSADSVSKDSGNI 299
Db 206 GKVLSSGYNFTTLPDPSGISVALED-----NTVQGVKNESGI-----KAGEI 249
QY 300 VLSAKEGEAEI-----GGVISQAQOAKGKLMITGDKVTLKTGAVIDLSGKEGETYL 353
Db 250 TLSAKGRKQALDSLVMNNGVLEATKYSNKNGKVLVSADNVLNNSNI-----KGEIVTF 304
QY 354 GGD--ERGEKNGIQIAKKTSLSEKGS-TINVSKEKGGFAIVGWDIALDGNINAQSGD 410
Db 305 GADVTNKKELKDKIKITSKTSKVTSPKINFTGKSVN-----INGNFGREDS-- 351
QY 411 IAKTGGFVETSGHDLFIKONAVDAKEWLLDF-DNVSINAEDPLFNFTINDEFFTGTGE 469
Db 352 -----TTHYKDEFKLTNEVN-----IDVPDENIRIAD-IEDNTGTG--ITGTGT 394
QY 470 ASDPKNSSELKTTLTNTTISNVLKNAWTNITASRKLTVNSSINI-----GNSHLILHSK 525
Db 395 SS-----FIQTGALSSLLANNGKVNKGN-VNLSGRHIDSPRGSDSLKLTKNK 443
QY 526 GORGGVQI-DGDIYSKGNLTIYS-GGMVDVHKNTITLDQGFNLITAAASVAPEGGNNKAR 583
Db 444 GH-----IDINNADIHSKGRLEFFITSLQNEEDFKSNITITDSKINLNGAMGL-GRSVDK 498
QY 584 DAANAQIVAOQVTTITGEKDFRANNVSLN-----GTGKGLNISSVNNLTHNLSTG 635
Db 499 DYDNRWQTEGS---QRKFDVKMSNVEFNQVDDVILAGGFEKVNLDKIV-----ATQG 549
QY 636 IN--ISGNITINQTTTRKNTSYWQTSKSHSNVLSALN-----LETGANFTFIKYI 682
Db 550 TNFYIDGGVSRN---GRKYEYGVLDLDRKQTLSELNQGRRRWGYDYDELDMNRYL---- 603
QY 683 SSNSKGLTQYVRSSAGVNFNGVNGMSFNLEKAGAKVNFKLPKNMNMNTSKPLPIRFLANI 742
Db 604 -----YR-----
QY 743 TATGGGVFFDIYAHNSGRGAELKMSKSEINISNGANFTLNS---HVRGD-----DAFKINKD 795
Db 606 -----PDLFATKNTGRSTIKDTTINTSN-SNIMKNGFVHLLAEKIKLONSKI--D 653
QY 796 LTINATNSFSLRQTKDDFDYGVARNAINSTYNISILGNVTL--GGQNSSSSTG----- 849
Db 654 ITFDKNSQDTLAQT-----NRLGMNGKVSIMSHIKIVGDEKIGISPTGYATM 703

```

QY 850 -----NITTEKAAVNTLEANNAP-----NOQNRDRVILKLG-----SLIVNGS 887
 Db 704 FLIGELIGBEKSLFVKGSHOGYFKTDGNTKIKAGKYSKEDLKTAINTGGRAAEVLLINGA 763
 QY 888 LSLTGENADIKG-NLTISSEATFKGKTRDTLNTGNTNNGTAENITQGVVKLGNTVDN 946
 Db 764 LGSADANDANIANMAFTIGDSANTK-TTIENADITALAPNGGTAYLSKDKDEIEVAPNSNF 822
 QY 947 GDLNITHAKRNORSIIGDI-INKKGSLNITDSNNDAEIQIGGNIQKSGEGLNLTSSDKI 1005
 Db 823 TFFELPREKLNLTNGKINGASTKLSERGFARLYDKINGVR---ASNLSABQLNVTDASEKI 879
 QY 1006 NITKQITIRKKGIDGEDSSSDATSN 1030
 Db 880 INTKLVS---SLDVEKLVSVAVCDA 901

RESULT 6
 ID OMPB_RICJA STANDARD; PRT: 1656 AA.
 AC 00653; 2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rompb)
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptidol).
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
 japonica."
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
 CC SIMILARITY).
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AB003681; BAA20138.1; .
 DR InterPro; IPR003858; rOmpB_rOmpB.
 DR Pfam; PF02708; rOmpA_rOmpB; 1.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1338
 FT CHAIN 1339 1656 120 KDA SURFACE-EXPOSED PROTEIN.
 FT DOMAIN 528 533 32 KDA BETA PEPTIDE.
 FT DOMAIN 1656 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 6.4%; Score 476.5; DB 1; Length 1656;

Best Local Similarity 21.8%; Pred. No. 1.9e-12; Indels 437; Gaps 78;

Matches 346; Conservative 204; Mismatches 596; Indels 437; Gaps 78;

QY 34 SEKPARMKVRHIALKPLSAMLISLG-----VTSIPOSVLASGLQGMVDVHGHTATMQVDG-- 87

Db 2 AOKPNFLK-----KLISAGLVLTASTATIVASFAGSANGAAIOQNRRTNGVAT--TVDGVG 54
 QY 88 -NKTIRNSV-----DAIINKKOFN-IDONEMV-----OFLQENNSAV-----FNR 127
 Db 55 FDOIVALANVAVAPNAVITANANGINLNTPAGSFENGFLFSNANNAVTVSEDTTLGFIN 114
 QY 128 VTSNOISOLKILDSNGOVFLINPENGIT-----IGKDAIINPN----- 165
 Db 115 NAANNANFNLTDR-GKTLITGOGIINVSOSAATHNAQINAVAKFENGGAALANNDLSGL 173
 QY 166 ---GFTASTL--DISNENIKARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVK 219
 Db 174 TIDFGAAASTLVFDLAN-----PTQKAPLIILADNALIVANGANGTLNTVNGFIQ 222
 QY 220 NEGVISVNGGSISLAGOKITISLIINPTITYSTIAAPENEAENVLDGIFAKGNGINVRAT 279
 Db 223 -----VSDKSFATV-----KAINIGD---GQGFMEFTN-AT 249
 QY 280 IRNOGKLSADSVS-----KDKSGNIVLSAKEGEAEFIGVISAQNOQAKGGKLMITGDKVT 334
 Db 250 NANALNLOAGGTTINFNGTDTGTRVLVLSKNGAATDFNVTGSLGNLKG-----IIE 301
 QY 335 LKTGAVI-DLSKGEGETYLVGGDERGEGKNGIOLAKKTSLEKGSTINVSGKEGKGFATVW 393
 Db 302 LNTVAINGOLIANAGPANAVIGTNNAGK---AAGFVVSVONGKAATIDGQVYAKDMVI- 357
 QY 394 GDIALIDGINAAGSGDIAKTGGFV-ETSGHDLFTKDNAIVDAKEMWLDGDFDNVSIADDP 452
 Db 358 -QSANANGOVNFRHIVDVGIDGTTAFKTAASIVAITQNSNFGT-----TDFGNLAAQVTP 412
 QY 453 LENVTGIDNFTGCEASDPKKNSELKTLTNTTISVYLNKNAWTMTNTASKLTVNSSI 512
 Db 413 ---DTWTLTGNF---TGDANNPONTAGVITFAANGILASASADA---NVAVTNNITAIERAS 464
 QY 513 NIG-----SNSHLILSHKGOORGGOVI-DGDITSKGNLTIYSGWVDVHKNTLDOGLFN 567
 Db 465 GGVGVOLSTHFAELRLGNAGSVFRLADGTVINGKVNQVILVGG-VLAAGATLD-GSAT 522
 QY 568 ITAASVAFEGGNKARDAANAKIVAQGVTTITGEGKDFRANNVSLNG-----TGKGLN 620
 Db 523 ITGD--IGNGGGGAALQSITLANDATKTLTLGG-----ANTISANGGTINQANGGTYK 574
 QY 621 ITSSVNNL-----THMLSGTINI----- 638
 Db 575 LTSTQNNIVVDCDLATATDQTGVVDASSLTNAQTTLTISGTIGIIGANNTLLOQFNIGSSK 634
 QY 639 -----SGNITINOTTRKNTSYWQTSKSHSNVNSALNLETGANFTFIKYIISNSKGLTQYR 694
 Db 635 TTLNGGNVAINELVIGNNGSVQFAHNTYLTITRTTNAAGOKIIFNPVNNNT-----T 687
 QY 695 SSAGVNF-NGVNGNMSFNL-KEGAKVNFKLKPNENMNTSKPLPIRELANITATGG--GSV 750
 Db 688 LAAGTNLGSAAANLAEINFGSKGARADTVLNVGEGVNL-----YATNITTTDANVGSF 740
 QY 751 FFDIYAHNSRGAEKLMSEINISNGANFTLNSHVGRDGAFFKINKDLTINATNSFSLROT 810
 Db 741 VF---NAGK-----NIVSGT-----VGGOGNKFNTVALDNGCTTVKF----- 775
 QY 811 KDDFYDGYARNALNSTYINISILGCVNLTGQNSSSSITGNITIEKAA----- 857
 Db 776 -----LGNATFN-----GNTTIAA-NSTLQISGNTADFIASADGCTGIVEFVNT 818
 QY 858 ---NVTLEANNAPNOQNIRDVRKLGSLVNGSLSTGENADIKGNITISE---SATFKG 911
 Db 819 GPINVTLNKOAVP-----VNALKQITVSG-----PGNVVNEIGNAGNYHG 859
 QY 912 KTRDTL-----NITGN-----FTNNGTAE-----INTQG 936
 Db 860 AMTDTIAFENSSLCGLVFLPSGIPFPNDAGNTIPLTIKSTVGTNETAEGFSVPVSGVDS 919
 QY 937 VVKLGNTVDGDLNITTHAKRNORSIIG-----GDIINKK-----GSLNITDSNDA 983

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OX Bordetella.
 RN NCBI_TaxID=520;
 RP [1]
 RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RA MEDLINE=90355839; PubMed=2388559;
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
 RT "Genetic characterization of Bordetella pertussis filamentous
 haemagglutinin: a protein processed from an unusually large
 precursor.";
 RT Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
 RL Mol. Microbiol. 4:787-800(1990).
 RN [2]
 RP SEQUENCE OF 1-3261 FROM N.A.
 RX MEDLINE=89202384; PubMed=2539596;
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
 RT "Filamentous haemagglutinin of Bordetella pertussis: nucleotide
 sequence and crucial role in adherence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
 CC -1- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
 CC INFECTION.
 CC -1- SUBCELLULAR LOCATION: SURFACE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M60351; AAA22974.1; .
 DR EMBL; M60351; AAA22975.1; ALT_INIT.
 DR EMBL; M60351; AAA22976.1; ALT_INIT.
 KW Antigen; Hemagglutinin.
 SQ SEQUENCE 3591 AA; 367420 MW; EF7418B3D06E5138 CRC64;
 Query Match 6.3%; Score 469.5; DB 1; Length 3591;
 Best Local Similarity 19.7%; Pred. No. 9 5e-12;
 Matches 381; Conservative 286; Mismatches 653; Indels 611; Gaps 88;
 QY 1 MN-KIYRLKSKRNALVAVSELARGDHST-----EKGSEKPAKMKVRIHIALKPLS 51
 DB 1 MNTLYRLVSHVGRMLVPVSE-----HCTVGNFTGCGTRGQAR-SGARATSLSVAPNA 53
 QY 52 -AMLLSLGVTSIPQSVLASLGQMDVHGATMQVDGKTIIRNSVD-----ALIHWKQ 104
 DB 54 LAWALMLACTGLPLVTHAQL-----VPGQTQVLQGGKVPVNIADPNSGGVSHNKFQ 109
 QY 105 FNIDQNEWOF-----LQENNN-----SAVFNRVTSNQISQLKGILDSNGQ 145
 DB 110 FNV-ANPGVVFNGLTDGVSIRIGALTKNPMLTRQASAILAEVTDTSPSRLAGTLEYVGK 168
 QY 146 ---VELINPGITIGDKAIINTNGFTAST-----LDISNENIKARNFTFEOT-- 189
 DB 169 GADLIANPGISVNGLSTUNASMLTJTGPRSVNGRIGLDVQQGVTVIERGCVNATGL 228
 QY 190 -----RDKALAEIV-----NHGL-----ITVGKDGSVNLIGGK 217
 DB 229 GYFDVVARVLKQGVSSKQKGLADIAVAGANRYDHTATRPATIAAGARGA----- 282
 QY 218 VKNEGVSIVNGGSSISLLAGOKIT-----ISDIINPTITYIAAPENEAVNLGD 265
 DB 283 ---AGAYADGTAGAMYGKHITLVSDSLGVRQLGSLSPS-----AITVSSQGEIATGD 336
 QY 266 IFAKGGINVRATIRNOGLSADSVSKSGNVLISAKEGEAEIGGVISAQNOQAK-GG 324
 DB 337 ATVQRPLSLKGVAVSAGKLAS-----GGGAVNVAGGAVKVIASASSVGNLAVQGG 369
 QY 325 KLMTIGDKVTLTKTGAVIDLSGKEGGETYLGGERGEKNGIQLAKKTSLEKGTINVS 384
 DB 390 KVOAT-----LLNAGGTVLLSGRAVQLGRASSRQALSVNAGGALKADKLSATRRVDGK 445
 QY 385 E-----KGGFAIWGDIALIDGNINAOGS-----GDIATGGGFVETSGHDL 425

DB 446 QAVALGSASSNALSVPAGGALKAGKLS-ATGRLDVDGKQAVTLGVSADGALSVSAGN 504
 QY 426 FIKDNAIVDA-----KEWLLD-----FDNVSINAEDPL-FNNTGINDEFTGTGEA-- 470
 DB 505 --RANELVSSAQLEVRGQREVALDDASSARGMTVAAGALAAARNLQSKGAIGVQGGEAVS 562
 QY 471 -SDPKKNSLKTLLNTTISNVLKNAWNTITASRKLTV-----NSSINIGNSHILHS 534
 DB 563 VANANSDAELRVGRGQVDLHLSAARGADISGEGRVNIIGRARSDDVKVSAHGALSIDS 622
 QY 525 KGQGG-GVQIDGDIITSGK-----GNLTIYSGGWV---DVHKN---ITLDQGLNI---T 569
 DB 623 MTALGAIGVQAGSVSAKDMRSRGAVTVSGGAVNLGVDQGVQVPRATSAGATVDRDVA 582
 QY 570 AASVAFEGGNKARDAANAKIV-AOGTVITITGEGKDFRANNVSLNGTGLNLISSVNL 628
 DB 683 AADLALQAG-----DALQAGFLKSAGAMTVNGR-----DAVRLDGAHAGGQL--RVSSD 729
 QY 629 THNLSGTINISGNITINOTTRKNTSYWOTSHDSHNVSALNLETGANFTFIKVISNSK- 687
 DB 730 GOAALGSLAAKGLTVS-AARAATVAELKSLD--NISV---TGERVSVQSVNSASRV 781
 QY 688 -----GLTTQYRSSAGVNFNGVNGKNSFNLKEGAKVFKLKPENMNT 730
 DB 782 AISAHGALDVGVKSAKSGIGLEGWAGVADSLGSDGALSVSGRDVAVRD----- 830
 QY 731 SKPLPIREFLANIT--ATGGSVFFDIYANHSGRGAEKMKSEINISNGANFTLNS----- 782
 DB 831 ----QARSLADISLGAEGGATL-----GAVEAAGSIDVRGGSTVAANSLSHANRD 875
 QY 783 -HVRGDDAFKI-----NKDLTINATNSFSLRQTKDDFYDGYARNAINSTVN 828
 DB 876 VRVSGKDAVRVTAATSGGLHVSQGRQLDLGAVQARGALAL---DGGAGVALQSKASGT 932
 QY 829 ISILGG-NVTLGGQNSSSI-----TGNITIEKA---ANVTLEA-----NNAPNOQN 871
 DB 933 LHVGGEHLDTGLTAAAGVAVDNGTGDVRVAKLVSDAGADLQAGRSMTLGIVDTTDLQA 992
 QY 872 IRDRVILKGLSVLNGSL-----SLTGNADIKGLNTLISEATFKGKTRDTLNT-GNFT 924
 DB 993 RAQKLELGSVKSDGGLQAAAGGALSALAAEAVAGALESG-----QGVTVDRASARAD 1048
 QY 925 NNGTAENITOGVVKLGNVTNDGDLNTHAKRNQSI-----IGGDIINKKGSINTD 978
 DB 1049 STGSVGI---GALKAGAVE-----AASPRRARRALRQDFFTPGSVVVRAQGNVTGR 1097
 QY 979 SNDAEIQIGNI--SOKEGNLTISSDKINITKOITTKKIDGEDSSSDATSNANLTIKT 1036
 DB 1098 GPHQGVLAQGDIIIMDAKGGTLLLRNDALTENGIVTI-----SADSAVLEHSTIES 1148
 QY 1037 K-----ELKLTEDLSISGFNKA-----EITAKDGR-----DLT 1064
 DB 1149 KISQSVLAAGKDKGKPAVSVKVAKKLFLNGTLRAVNDNNETMSGRQIDVDGRQITDAV 1208
 QY 1065 IGNSDNGSCAEAKTVFNNVKDSKISADGHVNTLNSKVKTSSNG-GRESNSDNDTGLT 1123
 DB 1209 TGEARKDES-----VVSDAALVADGGPIVVEAGELVSHAGGICNGRKNKENGASVT 1258
 QY 1124 ITAKNVENVNKTISL-----ITAKNVENVNKTISL-----EITAKDGR----- 1138
 DB 1259 VRTTGNLVNKGYSAGKQGVLEGGALTNFLVSGSDGTQRIEAOIRNRTGTFQSOAPAGT 1318
 QY 1139 -----KTVNITASEKVTTTAGS---TINATNGKA 1164
 DB 1319 AGALVVKAAEAIIVHDGVNATKMGEMQIAGKGGSPVTAVGAKATTSANKLSVDVASWONAG 1378
 QY 1165 SIITTKGDISGTISGNTV---SVSATVDLTKS-----GSKIEAKSGEANTVSTGT--- 1213
 DB 1379 SLDIKGGAQVTVAGRYAEHGEVSIQGDYTVSADAIALAAQVTOGGAANLTSHDRFRS 1438
 QY 1214 ----IGGTISGNTVNTANAGDLTVGNCAEINA-----TEGAATLTATGNTLTTEAGSSIT 1265

Db 1439 NKIRLMGLOVWAGPVSNTGNLKVREGVTVTAASFDNETGAEVMAKSAATLTT-SGAARN 1497

QY 1266 STKGQVLLAQNAGSAGSINAANVTNTGTLTTVAGSDIKATS-----GTL 1312

Db 1498 AGKMOV-----KEAATITVAASV--SNPGTFT--AGKDIITVTSRGFDNCKMESNKDI 1546

QY 1313 VINAK-----DAK--LNGDASGDSTE-----VNAVNASGSGSVTAATSSV 1351

Db 1547 VIKTEQFSGNRVLDAKHDLTVTASQADNRGSLKAGHDFTVQAQRIDNSGTMAAGHDATL 1606

QY 1352 NI-----TGDLNTVNGLNLIISK-----DGRN----- 1372

Db 1607 KAPHLRNTGVVAGHDHIIINSAKLENTGRVDARNDIALDVADFTNTGLYAEHDATLTL 1666

QY 1373 -----TVRLRGKEI--EVKYIOPG--VASVEEVIEAKRVLEKVK 1407

Db 1667 AOGTQRLVDQDHLPLVAEGTLRVKAKSLTTEIETGNPGSLIAEVOENIDNKQAIIVVGK 1726

QY 1408 DLS-----DEBRETAKLGVSARVEPNNTITVTQNETTRPSSQVLIISSEKACFSS 1461

Db 1727 DFLSSAHGVANEANALLWAA-----GELTVKQAN-----ITNKRAALTEA 1768

QY 1462 GNGARVCTNVA 1472

Db 1769 GGNARLTAAVA 1779

RESULT 9

HXAL_HAEIN STANDARD; PRT: 905 AA.

AC P44602;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization protein A).

DE HXAUA OR H10264.

GN Haemophilus influenzae.

OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBL_TaxID=727;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN-Bd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.B., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kariavaye A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Cotton M.D., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Sutton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.D., Geoghagen N.S.M., Gehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;

RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RT Science 269:496-512(1995).

RL FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.

CC SUBCELLULAR LOCATION: Secreted.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: U30712; AAC21929.1; -

DR TIGR: H10264; -

KW Transport; Signal; Repeat; Complete proteome.

FT SIGNAL 1 21 BY SIMILARITY.

FT CHAIN 22 905 HEME/HEMOPEXIN-BINDING PROTEIN.

SQ SEQUENCE 905 AA; 98833 MW; 2424013EB437A99D CRC64;

Query Match 6.3%; Score 466; DB 1; Length 905;

Best Local Similarity 23.7%; Pred. No. 2.5e-12;

Matches 251; Conservative 153; Mismatches 347; Indels 308; Gaps 54;

QY 72 QCMVYHGTATMQVDGK-TIIRNSVDALIIKQFNIDQNEWQFLOENNSAVENRVT 130

Db 28 QGSSVVVGEANVTIGKMTIDQKPTQIDHISFDIQCKEVEFPDANSVAYNRVT 87

QY 131 NQISQLKGIILDSQVFLINPNTIGTKDALINTNGTASTLDI-----SNEIKARNFT 186

Db 88 GNASQIOGKLTANGKVLPLANGVLIITOGAEINVAGLFAFKDLERISENGNGNKFT- 146

QY 187 EQTKDKAL---AEIVNHGLITVKGDSVNLIGGVKKEGVI-SVNGSISLLAQCK--IT 240

Db 147 RKLKDGQVVRKGOVINKGKIK-AKD-FVVLNGDKVINEGEIDATNGKVYLLSSGNTFT 204

QY 241 ISDIINPTITYSIAAPENAEVNLGDIFAKGNINVRAATIRNOKLSADSVDKSKNSV 300

Db 205 LSD-----SSISVALEDNAVQSI-----VQNEGII-----KAGDIT 235

QY 301 LSAKEGEAEI-----GGVISAQNOQAKGKIMITGDKVTLTKTGAVIDLDSKEGETYL 354

Db 236 LNAKGRNALDSLVNMNGVLEATKVSNGKGVLSADDDVQLNNKS--DIKGE--SEVVF- 290

QY 355 GDERGEGKNGIOLAKKTSLEKGS-TINVSKEGKGFAIWMGDIALIDGNIQAOGSDIAK 413

Db 291 ---TNEPKNIKITSQTSKVTSPKINFTGKSV-----NIN---GDFGR 328

QY 414 TGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSIINAEDPLENNTGINDPEFTGTGEASDP 473

Db 329 D-----DSRAH--YNEHKKRLDTE-----VNIQVDPDENIRIAEKDNTGTGTGD- 371

QY 474 KNSCLKTLTNTTISNLYKNWNTNITASRLKTVNSNINI-----GNSHLILHSGKORG 529

Db 372 -----SFIQTGALSLLANNKYNL-KGDVNIISGRHIDSFRGSDSLKLNTQGH- 421

QY 530 GGVQID-GDITSKGGNLTYS--GGWVHKHTITLQGFNLITAAVAF----- 575

Db 422 --IKINHADIHSTGRLEFFITSLQNEKDSQSDITITDSKINLNGANGWGLRSLDKRNCNQ 479

QY 576 -----EGGNKARDAANAKIVAQ--GTVTITGEGKDFRANVNSLNGTGKGLNIIS--SV 625

Db 480 RWCRTETSORKKFVHMNRNVFDDQVDVVVAGGFKVNLNDIV--ATGKTNYIDGGVSR 537

QY 626 NNLTHNLSTINISGNTINQTRK-----NTSYW-----QTS 658

Db 538 NNSRVEY-CVLDLDRKTLSELQDQRRRWKYVNDLDMNKAYWHRFDMFATKNTGRSTI 596

QY 659 HDSHNVV--SALNLETG-----ANFTFKYISSNSKGLTTOYRSSAGVNF 701

Db 597 KDTENISNKLNGFVHLLAEIKLDSKIDITFDK---DNSQDISTSQ-----INR 647

QY 702 NGVNGKNSFLREGAKVNFELK---PNEPMNTSKPLPIREFLANITATGGSVFFEDIYAN 757

Db 648 LGNNGKYSM-----VNSHIKIVGDEKSDISAKAPYATMFLIGELIGESKSIIF--VKS 697

QY 758 HSGRGAEKMSINISGANFNLNSHVGRDGAFAKINKOLTINATNSFSLROTDDFDYDG 817

Db 698 HQSY-----TFRTDGTAKIAGKNSKDLKIT----- 723

QY 818 YARNAINSTVNTSILGNTVTLGONSSSITNITEKAANYTLEANNP-----NOONIR 873

Db 724 ----ALNT-----GGR-----TGKEVI-----INGAPGSIDNDANTA 751

QY 874 DRVIKLSLLVNGSLSLUTGENADIKONLTISEATPKGTRDPTLNTGFTNNGTAETNI 933

Db 752 NMAFTIGD---NANTKTIIENAD-----ITALAPNGGGATYLS- 785

QY 934 TQGV-VKLGNTVNDGDLNITTHAKRNQORSTIGGDI-INKKGSINITDSNNDAEIQIGGI 991

Db 786 SKGVEIEVNPNSNFTFELPREKNFNQTKIGDSTKLSERGFARLYDKINGVR---ASNL 842

Qy	992	SQKEGNLTISSDKINITKQITIKKGIDCEDSSSDATSNA	1030
		: : : : : : : : : :	
Db	843	SAQLNVTDASEKIINTKLVS---SLDVEKLVSVACDA	878

RESULT 10

```

YDEK_ECOLI
ID YDEK_ECOLI STANDARD; PRT; 1325 AA.
AC P32051; P76140; P77168;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein ydek precursor (ORFT).
DE YDEK OR ORFT OR BL510.
GN YDEK OR ORFT OR BL510.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Pernthaler J., Broun P.,
RR Riley M., Collado-Vides J., Glasner J.D., Rode C.K.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.C.,
RR Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Asai H., Baba T., Fujita K., Hayashi K., Inada T.,
RR Kaba H., Kashimoto K., Kimura S., Kitakawa M., Kikuchi
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RR Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Ogasawara
RA Nakade S., Seki Y., Sivasubramam S., Tagami H., Teraoka
RR Sempel G., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi
RA "A 570-kb DNA sequence of the Escherichia coli K-12
RR corresponding to the 28.0-40.1 min region on the lambda
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 595-1325 FROM N.A.
RC MEDLINE=94100243; PubMed=8274505;
RX Cartwright P.J., Timms M.W., Lithgow T., Hoef P.B.;
RA "An Escherichia coli gene showing a potential analogue
RR to the genes for the mitochondrial import site protein
RA MOM38.";
RR Blochim, Biophys. Acta 1153:345-347(1993).
RL -1- SUBCELLULAR LOCATION: Attached to the membrane (Potential).
RN [4]
RP -1- SIMILARITY: TO E.COLI YFAL.
RX -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT
RA ISP42 AND REF.38.
RR -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN
RA FRAMESHIFT IN POSITION 653.
RR -----
RL This SWISS-PROT entry is copyright. It is produced
RN between the Swiss Institute of Bioinformatics and
RX the European Bioinformatics Institute. There are no
RA use by non-profit institutions as long as its con-
RR tent is not modified and this statement is not removed. Usage
RA of this entry requires a license agreement (See http://www.ebi.ac.uk/
RR or send an email to license@sib.slb.ch).
RL -----
RN EMBL; AE000246; AAC74583.1; -
RX EMBL; D90793; BAAL1590.1; ALT INIT.
RA EMBL; D90794; BAAL1597.1; ALT INIT.
RR EMBL; X73295; CAAS1730.1; ALT_FRAME.
RA PIR; S34315; S34315.
RR Ecogene; EG11780; YDEK.
RL PROSITE; P500013; PROKAR_LIPOPROTEIN; 1.

```

[illegible]

```

QY 840 GONSSSINGNITIEKAAVNTLEA-----NAPNOONIRDRVILKGLSLVN-GSLSLTGEN 894
DB 684 GQVDMMLITGD---DKGKVIHDGHTVFNAGNTYS-----GKTLVNDGLLTIASHT 732
QY 895 ADIKGMLTTSSESATFKGKTRDL---NITGNFT-----NNGTAENITQ----- 935
DB 733 ADGVTGMSSEVITIANPGTLDILASTNSAGDYTLNALKGGDLRMVQLVSSDKMGCFTHA 792
QY 936 -----GVVKLGNTVDGDLNTT---THAKRNORSITIGGDIINKKGSUNITDSNDAEIQ 986
DB 793 TGTEFAGVAQLKDSFTFLERDNTAALTHAMLQ-----SDSENTTSVK 834
QY 987 IG-----GNISQKRGNTISSDKINIKITIKKIDGEDSSSDAYSANMLTIKTKELKLT 1042
DB 835 VGSQSIGGLAMNGGTIIFDTD---IPAATLAEGYISVDTLVVGAGDYTWKGRNYOVNGT 890
QY 1043 EDLSISGFNKAETAKDGRD-----LITGNSNDGNSGAE---AKTV-----TFN 1083
DB 891 GDVLID-----VPRPNDPMANPLTTLNLEHDDSHVGVLVKAQTVIGSGSLTLR 943
QY 1084 NVKDSKISADGHNVTLSNKKVKTSSSNGGRESNDNDTGLTITAKN-VEVKNKDIITSKTVN 1142
DB 944 DLQGDVEAD-----KTLHIAQNGTVVAEGDYFRLLTAPGNGLYVN---YGLKALN 992
QY 1143 ITASEKVTITTAGTINATNCKASITTKTGDISGTVISGNTVSVSATVDLTKGSKIEAKS 1202
DB 993 IHGGOKLTL-----AEHGGA--YGATADMSAKIGGEG-----DLAINTVROVSLSN 1036
QY 1203 GEANVTSATGTIGTISGNTVNTVANAGDLTVGNGA--EINATEGAATLTATGNTLTTEA 1260
DB 1037 GONDYQAGTVQMGTLRTDADGALGNTRELNISNAIVDLNGS-----TQVETFTGQM 1090
QY 1261 GSSITSTKGQVDLLAQAQNGSTAGSINAANVTNTTGLTVVAGSDIKATSCITLVINAKDAK 1320
DB 1091 GSTVLVFEK-----ALTVNKGGI---SQGLT---GGGNLNVGTGTLAIEGLNAR 1134
QY 1321 LNGDAS--GDSTEVNNAVNSGSGVTAATSSSV---NITGDL-NTVNGLNLIISKDGRNTVR 1375
DB 1135 YNALTSISPAEVSLEDNTOGILGRGNIANDGLLTLKNVTGEIRNSISGKIVSATATDVE 1194
QY 1376 LRG 1378
DB 1195 LDG 1197

RESULT 11
OMP_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OmpB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor."
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;

```

MEDLINE=90136087; PubMed=2515418;
 Gilmore R.D. Jr., Joste N., McDonald G.A.;
 "Cloning, expression and sequence analysis of the gene encoding the
 120 kD surface-exposed protein of Rickettsia rickettsii.";
 Mol. Microbiol. 3:1579-1586(1989).
 -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 LAYER WITH HEXAGONAL SYMMETRY.
 -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: X16353; CAA34403.1; -
 InterPro: IPR003858; rOmpA_rOmpB.
 Pfam: PF02708; rOmpA_rOmpB; 1.
 Antigen: S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
 CHAIN 1 1333 32 KDA BETA PEPTIDE.
 FT CHAIN 1334 1654
 FT DOMAIN 1181 1188 POLY-THR.
 FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
 SQ

Query Match 5.8%; Score 429.5; DB 1; Length 1654;
 Best Local Similarity 22.1%; Pred. NO. 1.6e-10;
 Matches 357; Conservative 179; Mismatches 583; Gaps 83;

```

QY 34 SEKPARMKVRHLALKPLSAMLISLG-----VTSIPQSVLASGLQGMVHVHGTATMOVDG-- 87
DB 2 ACKPNFLK-----KLISAGLVLTASTATVASFAGSAMGAATQQRNRTNGAAT-IVDGAG 54
QY 88 -NKTITRNSVDATINWKQFNIDQNMVQFLOENNNNSAVFNVRVTSNQISOLKGILDSNGV 146
DB 55 FDOTAAPANGVALN-AVITANANGINF---NTPAGSFNGLLLTANNL-----TIT 136
QY 147 FLINPNGITIGKDAIINTNGFTASTLDISNENIKARNFTFEOTKDKALAEIVNHGLITVG 206
DB 101 -----AVTSED---TTLGF-----INVVHNSHFNLTNAGKTL-----TIT 136
QY 207 KGSVNLGGKVKNEGVISV---NG-----GSISLLAGOKITISDIINPTITS 252
DB 137 GCGVTNAQAATKNAQNVVVFQFNNGAAIDNNDLKGVRIDFGAPASTLVFNLANPT---- 192
QY 253 IAAPENEAVNLGD--IPAKG--GNINVRATITRNQKLSADSYSKDSKSNIVLSAKEGA 308
DB 193 ---TQKAPILGDNAVIANGVNGLNVTNGFI---QVSNKSFATVVKAINI-----A 237
QY 309 EIGGVI-----SAQNOQAKGGKLMITGDKVTLTKTGAVIDLSGKEGETY-----LGG 355
DB 238 DGQGIIFNTDANNANTLNLAQAGTTINFTG---TDGTGRVLVLLSKHAAATNFTITGSLGG 294
QY 356 DERGEKG-NGIOLAKKTSLEKSGSTINVSKEKG-----GFAIVW--GDIALIDGNI----- 403
DB 295 NLKGVIEFNTVAVDQGLTANAGANAIVGTNNGAGRAAGFVSVVDNGKVATIDGOVIYAKD 354
QY 404 -----NAQSGD---IAKTGGFVETSGHDLFIKDAIYDADAKEWLLDFDNVSNABDPLF 454
DB 355 MVIQSANATGOVNFRLHVDVGADGTTAEKTAASKVITITQDSNFGNTDFGNLAAQIKVP-- 412
QY 455 NNTGINDPEPTGTGEASDPKKNSELKTLTNTTISNVLKNAWMTNTTASRKLTVNSINI 514
DB 413 NATLTGTF---TGDASNPCNTAGVITFDANGTLESASADA---NVATNNTTIAEASGA 466
QY 515 G-----SNSHLILHSGKQGG-----GGVQIDGD 537
DB 467 GVVOLSGTHAAELRLNAGSIFKLADGTVINGKVNQATALVGGALAACTITLDGSATITGD 526

```

```
QY 538 ITSKGG-----NLTIVSGGVVHKKITLDOGFNLITAA-----SVAFEGGNKKAADAAN 587
D 527 IGNAGGAALQRTILAN-----DAKKTLLT--GGANIGAGGGTIDLQANGTKIKLTSTQN 580
QY 588 AKIV-----AAGVTITGEGKDFRANNSVNGTGLNLTSSVNV 626
D 581 NIVVDFDLATADOTGVWDASSITNAQ--TITINGKITGTGANNKTLG-----QFNIGS--- 632
QY 627 NLTHNLSGTTINISGNTTINOTTRKNTSYWQTSHDHSHNVSNALNLETGANFTIKVSSNS 686
D 633 -----SKTVLSNGVAINELVIGNDGAVQFAHDTVLIITTTNAAGOGKLIENPVVNGT 686
QY 687 -----KGLTQYRSSAGVNF--NGVNGNSFNLKEGAKVNFKLKPNENMNTSKPLPIREL 739
D 687 TLAAGTNLGSATNPLAEINFGSKGVNDVTLNVGCVNL-----YA 727
QY 740 ANITATGG--GSVFEDIYAHSGRAELKMSINISNGANFTLNLSHVGRDDAFKINKDLT 797
D 728 TTTITTDANVGSFVN-----AGGTNI--VSGTVGGQGNKFNITVAL 767
QY 798 INATNSFSLRQTKDDFYDGYARNAINSTYNIISILGGNVTLGGQNSSSI-----TGN 850
D 768 ENGTTVKFLGNAT-----FNGNTTIAANSTLQI---GGNTADCVASADGTGIVEFVNTGP 820
QY 851 ITIEKAANVTLEANNAPNOQNRDRVIKGLSLLVNGSLSLTGENADIKGNLTISE---SA 907
D 821 IT-----VTLNKQARP-----VNALKQITVSG-----PGNVVINEIGNAG 855
QY 908 TFKGTRDTL-----NITGN-----FTNNGTAETINITQGVVVL 940
D 856 NHGAVTDTIAPENSSLSGAVVFLPRGIPFENDAGNMTPLIKSTVGNKTAKGDFVPSVVVL 915
QY 941 G--NVTNDGDLNITHAKRNORSIIG-----GDIINKK-----GSLNTDSNDAEI 985
D 916 GVDVIAQGV-----TGDONNVGLGSDNGIIVNATLLYAGISTLN---NNOGTV 965
QY 986 QIGGNISQKGNL-----TISDKINITHQITIKKID--CEDSSSDATSNANLTIKTKE 1038
D 966 TILSGVPNTPGTVGLGTGIGASKF---KQVTFDYNNLGNIIATNATINDGVTVITG- 1021
QY 1039 LKLTEDLSISGNKAEITAKDGRDLTIGNSDNGSGAEAKTFTNNVSKDSKISADGHNT 1098
D 1022 -----GIAGI-----GPDGK--ITLGSVN--GN-----GNVR-----FADGILSN 1052
QY 1099 LNSKVTSSSNGG-----RESNDDNTGLTITAKNVENKDI 1135
D 1053 STSMIGTKANNQTVTLGNAFVGNIGSDTPVASVFTGSDSGAGL-----OGNIYSQV 1107
QY 1136 TSLKTVNITASEKVTITAGSTINATNGKASITTKTGD--SGTIS--GNTVSVSATVDLTK 1193
D 1108 IDFGTYNLGIVNSNIILGGGT--TAINGKIDLVNTLTFASTGTSTWGNNTSIETTLTANG 1166
QY 1194 SGSKIARSGEANTVTSATGTIGGTISGNTVNTANA-----GDLT---VNGGAEINATEG 1245
D 1167 NIGHVILEGAQVNTTGT-----TTIKVODNANANFSGTQTYTLTQGGARFNGTLG 1219
QY 1246 AATLTATGN-----TLTTEAGSITSTKGQVLLAONGSIAGSIANAANVTLLTGTLLT 1299
D 1220 SPNFAVTGSRNFVNSLIRAAQDYVITR-----TNAENVVTN-----D 1259
QY 1300 VAGSDIKATSG-----TLVINAKDALNGD---ASGDSTEVNNAVNSGSGSVTAATSSV 1351
D 1260 IANPPGAGPVQONVTFVFNATNTAAYNNLLAKNSANSANFVGAIVTDSAAITNVQL 1319
QY 1352 NITGDL-----NTVNGLINIISDKGRNTRVLRGKEIEVKYIQCVASVEEVIE 1398
D 1320 DLAKIDIAQIAGNLGALRYLGTGP--ETAEWAGPEAGA--ISAAVAGDEAID 1367

RESULT 12
BIGA_SALTY
ID BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; O9XCQ3;
```

```
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein bigA precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028;
RA Stojiljkovic I., Valentine P., Heffron F.;
RT "Salmonella typhimurium rls homolog.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91100301; PubMed=19871123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cys6 plasmid to overcome limiting shrohome
RT cofactor.";
RL J. Bacteriol. 173:325-333(1991).
CC -! CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF133696; AAD39458.1; -
DR EMBL; AF008859; AAL22340.1; -
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StyGene; SG10437; bigA.
KW Virulence; Repeat; Signal; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 1953
FT POTENTIAL.
FT PUTATIVE SURFACE-EXPOSED VIRULENCE
FT PROTEIN BIGA.
FT 15 X 11 AA TANDEM REPEATS.
FT REPEAT 101 252
FT REPEAT 101 103
FT REPEAT 104 113
FT REPEAT 114 122
FT REPEAT 123 133
FT REPEAT 123 133
FT REPEAT 134 144
FT REPEAT 145 155
FT REPEAT 156 166
FT REPEAT 167 177
FT REPEAT 178 188
FT REPEAT 189 199
FT REPEAT 200 210
FT REPEAT 211 221
FT REPEAT 222 232
FT REPEAT 233 243
```

FT REPEAT 244 252 15 (INCOMPLETE).
 FT CONFLICT 207 D -> DRGDDVTPDD (IN REF. 1).
 FT CONFLICT 514 A -> R (IN REF. 3).
 FT CONFLICT 1698 D -> N (IN REF. 1).
 FT CONFLICT 1798 QYLE -> ITIQ (IN REF. 1).
 FT CONFLICT 1836 SA -> T (IN REF. 1).
 SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 5.8%; Score 429.5; DB 1; Length 1953;
 Best Local Similarity 20.4%; Pred. No. 2e-10;
 Matches 336; Conservative 225; Mismatches 581; Indels 503; Gaps 79;

QY 73 GMDVHGTATQVGDGKTIIRNSVDAILNKKOFNIDONEMVQFLOENNSAVFNRTSNQ 132
 DB 310 GWDV-----TDANAALIEGTQENGLYWKY-----DSRGYLLIADDDNTVISG--DDQ 354
 QY 133 ISQLKILDSNGOVFLINPENGITIGKDAIINTNGFTASTLD-----ISNENIKARNFTF 186
 DB 355 AHNSDRGMDISGQ---DRTGVIIISGDRVNTLTGSSVTDGATGMVISGDG--TNTTISG 409
 QY 187 EQTKDALAEIVNHGLITVKGDSVNLIG-----KVKNEGVISVNGSGSISLLA 235
 DB 410 HSTVDNATGALISNGTTFNFGDIAVSGGTAIIIDGNATIKTGTSDISG-----A 463
 QY 236 GQKTIISDIINPTTYSIAAPENAEVNLGDFAKGGINVRAATIRNOKLSADSVSKDK 295
 DB 464 GSTGTVIDGNARVN-----NDGDMTITDGTG-----GHITGDNVVIDN 503
 QY 296 SGNIVLSAEG-----EAEIGGVISAQOQAKGKL--MITGKV--TLKGTGAVIDLSGKE 347
 DB 504 AGSTTVSGADATALYIEGDNALVINEGNOTISGAVGTRIDGDDAHTTNGD--IAVDGAG 562
 QY 348 GGETVLGGDERGEKNGIQIAKTSLEKGSTINVSKEGGEFAIVMGDIALIDG--NINA 405
 DB 563 SAAVIINGD-----NG-----SLTQA-----QVLLDTGAMGIIT 592
 QY 406 QGSGDIAGTGFVETSGHDLFIKDNAIVDAKEWLLDFDNYS--INAEEDPLFNNTGINDEF 463
 DB 593 YGTNEAKNTG-----NATV-----RDADSVGVVAGEKNTFKNKGDIDVS 633
 QY 464 PTGGEASDPKKNSELKTTLTNTTISYLNKAWMTNITASRKLTVNSINIGNSHLIH 523
 DB 634 LNTGT-----ALVSGDMS-----QVLLDGDINWVS-----VQD 661
 QY 524 SKG--QRGGGVQIDGDTISKGGNTLIYSGWVDV--HKNITLDQGFNLNITAAVAPE-- 576
 DB 662 SEGVSATGVSVSGDSNA-----VDITGVNIVSADYGQDDLAAGAPPLTGVV 709
 QY 577 -GGNNKARDAANAKIVAGQTVITIEGKDFRANVSVLNGTKGLNLISSVNNLTHN---L 632
 DB 710 VGGNGNTVTLNGLNIDDDNSATG--GYLDVGLSVTGDDNDVEIDGGI--NITHSEDPL 767
 QY 633 SGT-----INISGN--ITINQTRKNTSYQWTSH-----DSHWYSAL 668
 DB 768 DGTSDIATIGSVSGNSTVTLNGHSTIDTNTVVGHVILARVNNCGSILGDDSVVDVNVVS 827
 QY 669 NLETGANFTFKYIISNSSKGLTTOYRSSAGVNFNGV-----NGNMSFN----- 711
 DB 828 YIPTG--YTYTNALLMADGEG--TSIENKGDITSHGVSVIRADNGSEVNSGDIILVYATS 884
 QY 712 -----LKEGAKVNFKLPKNENMNTSKPLPIRFLNATITATGGGSV----- 750
 DB 885 SNSSEDRAAITRASGEGSAVHNKAGGD-----TTLISDQTPOGGGIEVYPLKWY 934
 QY 751 ---FFDIYVNHSGRAELKMSBEINSGANFTLNHVRGDDAFKI--NKDITINATNSNF 805
 DB 935 THTFYAMASDYG-----DVVNEGATI--HLQAGAGVYGVTSARGLNENGYL 982
 QY 806 -SLRQTKDDFYDGYARNAINST-----YNISILGNNVTLGGONSSSSITGNITIE 854
 DB 983 DGLVPLDD-----ENNTITSYQWPSLYLTSSGMVAGSTADAG--DATAINTGNITV- 1034
 QY 855 KAANTYLEANAPNOQNTDRVILKUGSLVNGSLTGTENADIKNLNLTISESAFKGKTR 914

DB 1035 -----NNAG-----FGMMALNGGTAINQGVITLAD--DGVTG 1065
 QY 915 DTLNITGNFTNNGTAENITQGVVKL-----GNVTNDGDLNITTHAKRNORS 961
 DB 1066 QADELVGAALNGVWINDTSGVINIDADYQOAFLSSSSYIINNGSINLNGSPMDTDS 1125
 QY 962 IIGGDIINK-----KGSNLNTDSNDAEIOIGNISOKEGNLTISS----- 1003
 DB 1126 HMGTPPTDKIWIQSLPGSGSDTRTSDFPTAGTLA--NYGTETLNGDVVDVNGWLNIA 1184
 QY 1004 --KINITKQITIKKIGDEDS-----SSDATSNANLTIKTKELKLTED--LSISGNKAEIT 1056
 DB 1185 GASLTVNGTIVTINGANALANYGLDADAISTWHSFLNEADGSITDILLUNG----- 1237
 QY 1057 AKDRDLTIGNSND--GNSGAEAKTVTFFNNYKDSKISADGHNVTLNSKVTSSSNGGRES 1114
 DB 1238 -----DVTFFYNGDFTGSIAGTSYQOEIVNTGDMTVAEDGKSLV-----SGSFYF 1282
 QY 1115 NSDNDTGTITAKNVFNKDIITSLKTVNIT--ASEKVTTTAGSTINATNGKASITTKTGD- 1172
 DB 1283 YNEEDATLTNSGSAVEGGEN---TIINLTRANDSLTVQVNSGTTITATNGYSAITVNGSN 1338
 QY 1173 -----ISGTISG-----NTVSVA-----TVDLTT 1192
 DB 1339 DPKWIWNTATGVINGINPDAPLINLGRGYNFGNOGTINVOGDNAVAISGGTSSVYINLVN 1398
 QY 1193 KSGSKIEAKSGEANTVSATGTIGGTISGN--TVNVVTANA-----GDLT--VGN 1236
 DB 1399 SGTINVTGEOGEDGTNGTGLIGIKGNGNATINNTADGVINVYADDSVAEGKTKAIIN 1458
 QY 1237 GAEIN--ATEGAATLTATGNTLTTEAGSSITSTKGOVDLLAONGSTA--GSIANAANVTLN 1292
 DB 1459 NGEINLLCDSGCDITAPGTTGTQ-----NDHNGTADIVIPATTPATPGSITPPADPN 1512
 QY 1293 TTGTLT--VAGSDIKATSGTLVINAKDLNGDASGDSTEVNVAASGSV-----AA 1346
 DB 1513 APOQLSNYIVGVNAGSSGTL-----KANLVIGDNVYKVDITGFTSGTADTVVVDNAF 1565
 QY 1347 TSSV-----NITGDLTVNGLNIISKDGNVRLRKKEIEVKYIQGVASVEVEIAKRV 1402
 DB 1566 TGSNIOGADNITSTSVVNAQSQDAGNVDVTM-----TRNAVADVATDSVSDVAQA 1619
 QY 1403 LEKVKDLSDEERETLAKGV-----SAVFEVFPNNTITVNTQ-----NEFTTPSSQVII 1452
 DB 1620 LD--AGYTNNELVTSUNVGTATLNSALKOVSGAQATTVFREARVLSNRFMTLADAAPOI 1677
 QY 1453 SEGKACFFSSGNGARVCTNVADGQP 1477
 DB 1678 KDGLA-----FNVYAKGDP 1691

RESULT 13

OMP_B_RICCN STANDARD; PRT; 1655 AA.
 ID OMPB_RICCN
 AC O9KKA3; O9KX98; O9XC45;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rompB)
 DE antigen) (120 kDa outer membrane protein) (Surface protein
 DE OMPB OR RC1085.
 GN OMPB OR RC1085.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_TaxID=781;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,

Db 1045 RFVDG-ILSHSTSMITTKANNNGTVYLGNAFVNGTGDSD-----TPVASVFTGSD 1095
QY 1110 GGRESNDNDTGLTITAKNEVNDKITSKVTNITASEKVTITAGSTINATNGKASITTK 1169
Db 1096 GG-----AGL-----OGNIYSQVIDEFTYNLGISNSVILGGOT-TAINGKINLRN 1141
QY 1170 TGDIT-SGTIS-GNTVSATFVDLTTKS-GSKIEAKSGEANVTISATGCTIGCTSGNTVNT 1226
Db 1142 TLTFASGTSTWGNSTSTETTLTLANGNIGNIVILEGAQVNAIT-TGT-----TTIKVQ 1193
QY 1227 ANA-----GDLT---VNGAEINATGAATLTATGNTLTTEAGSSITSTKGOVDLLAONG 1278
Db 1194 DNANFSGTGTTLTIGGARFNGTLGGPNFVT-----GSRFVNGVGLIRANQDY 1245
QY 1279 SIAGSINAANVTNTTGLTTFVAGSDIKATSG-----TLVINAKDAKNGD---ASGDST 1330
Db 1246 VIETNNAENVIN-----DIANSFPGGAPGVQNVTFVNATNTAAYNNLLAKNSAN 1299
QY 1331 EVNAVASGSGSVTAATSSVNTIGDL-----NTVNGLNIIISKDGRNTVRLRGKETEVKY 1385
Db 1300 SANFVGAIIVTDSNAITQAIDVAKDIOAQLGNRLGALRYLGT-ETAEAGPEAGA-- 1355
QY 1386 IQPGVASVEVIE 1398
Db 1356 IPAAVAAGDEAVD 1368

RESULT 14
HLVA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN SHLA
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia
OX NCBI_Taxid=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC STRAIN=SN8;7037; PubMed=3290200;
RA MEDLINE=86257037; Braun V.;
RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia
marcescens."
RL J. Bacteriol. 170:3177-3188(1988).
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
DEFINED.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
REQUIRES SHLB FUNCTION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M22618; AAA50323.1; -;
DR PIR; A28182;
KW Hemolysin; Toxin; Outer membrane; Signal.
FT SIGNAL 1 30
FT CHAIN 31 1608 HEMOLYSIN.
SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 5.7%: Score 425; DB 1; Length 1608;
Best Local Similarity 21.0%; Pred. No. 2, 4e-10;
Matches 351; Conservative 216; Mismatches 576; Indels 528; Gaps 79;
QY 2 NKIVRLKSKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALRPLSLAMLSLGVTS 61
Db 3 NNPFRLSAGKLAALAILAASAGAYAE-----IVAANGANG 41
QY 62 IPQSVASLGLOMDVHGTATQVDGDKNTIIRNSVDAILNKQFNIDQ----- 109
Db 42 PGVSTAATGAOVVDIV-----APNGN-GLSHN-----QYQDFNVNPGAVLNSREAG 88
QY 110 -NEMVOFLOENNN-----SAVFNRTSNQISOLKG-----ILDSNGOVFLIIPNGITGKD 159
Db 89 LSQLAGQLGANPLNGREASVILNEVGRNPSLLHGOQOEIFGMAADYVLANPNGISCOSC 148
QY 160 AIINT-----NGFTASTLDISNENIKARNFT----- 185
Db 149 GFINTSHSVLVGNPLVNGVLQGYSTFCNRNTLSLNGTLNAGGVLDLIAPIKIDSRGEVI 208
QY 186 ---FQTKDKALAEIVN---HGLITVKGDSVNLIGGKVKNEGVISVNGSGISLLAGOKIT 240
Db 209 VDDEKOSNGKVTSAAINAISGLNRVARQGV-----QASQQMPTALDSYILGSMOAGRIN 263
QY 241 ISDIINPTITYSIAAPENEAENVLDGIDFAKGGNINVRAATIRNOGLKLSADSVSKDKSGNIV 300
Db 264 I-----INTAOGSGVKLAGSLNAGDELKVKAYDIRSESRV--DDASSNKGNDN 310
QY 301 LSAKEGEAEIGVISNAQ---NOQAKGGKLM-----ITGDKVTLTKTGAVIDLS 344
Db 311 YQNYRGYVNDRSSQTLRTTELKGNISLVADNHAHTATDIRGEDITLQ-GGKILFD 369
QY 345 GKEGGTYLGGDERG---EGKNGIQAKTSLEKSGSTINVSNGKEGGAIFWGDIALIDGN 402
Db 370 GOOLKOTGHTDDRWFYSWOYDVTRELOLQAGSTVAASGSAK---LISTQEDVKLLGAN 427
QY 403 INAOGSGDIAKTGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGIND 462
Db 428 VSADRALSV-KAARDVHLAG--LVEKD----- 451
QY 463 FPTGTGEASDPKKNSLKTTLTNTTISNYLKNWMTNITASRKLTVNSNINIG--SNSHL 520
Db 452 -----KSSE-----RGYQRNH-----TSSLRTGRWNSD- 475
QY 521 ILHSGORGGOVDIGDITSGKGNLTISYGGWYDVHKNITLD-----QGFLNITAAVAF 575
Db 476 --ESELKASELSRSEGLTLKAGRNVSOTQAKVHAQRDLTIDADNIOIQGVQKATANAKAV 533
QY 576 E-----GNNK-----ARDAANA-KIVAOGT-----VTITGEKDFRANNVSL 612
Db 534 RDDKTSWGGIGGDNKNNSNRREISHASELTSGTLRLNGOQGVITIG-SKARGQKGEV 592
QY 613 NGTGKGLNIISVNNLTHNLSGTINISGNITINOTTRKNTSYWOTSHDSHWNSALNLET 672
Db 593 TATHGGURIDNALSTTVDKIDARTGAFNIT--SSSHKADNSYQSS-----TASLKS 643
QY 673 GANFTFTKYISSNSKGLTTQYRSAGVNFNGVNGNMSFNLEKAGKVNFKLPKNENMTSK 732
Db 644 DTNLTLYSHKDADVIG--SQVAGGELSVESEKGTNI--NVKAAER-----QONIDEQK 692
QY 733 PLPIRFLANITATGGGVSVFFDIYANHSQ-----RGAELKMSSEINISNGANFTLNSHVR--G 786
Db 693 -----TALTUNG-----YAKEAGDKOYRAGLRIEHTRDSEKTRTTRTENSASSLSGG 737
QY 787 DDAFKINKDLTIN-----ATNSNFSLRQTKDDFYDVGARNAINSTYINISILGGNVLGGQ 841
Db 738 SVKLKAEKDVTFSSKLVADKGDASVSGNVSFLAADDKKTASN-TEOTKIGGGFYITGG- 795
QY 842 NSSSITGNITIEKAAANVTLEA---NNAPNOONIRDRVILKGLSVLNGSLTGENADIK 898
Db 796 -----IDLGLS--GVERGYENKTKOQS--SKAITSGS-----DVK 827
QY 899 GNLTISESATPKGKTRDTLNTITGNTNNGTAEINITOGVVKGLNVTNDGDLNITTHAKRN 958

```

Db 828 GNLIIN-----ARDKLTOGAHQHVGAGYQENAGVDHLA-----AADTASTTTTKTD 875
QY 959 QRSIIIGDI-----INKGSLNITDNNDAEIIQIG-----NISQKEGNTLISS 1002
Db 876 VGVNIGANVDYSAVTRPVERAVCAAKLDATGVIND-----IGGIGAPNVGLDIGAGGSS 931
QY 1003 DKNITKQ-----ITTKGIDGEDSSDTSNA-NLTIKTKELKLTEDLSISG 1049
Db 932 EKSSSSQAVSVVQAGSIDINAKGEVRQGTQYQASKGAVNLTDASHR-----SEAA 984
QY 1050 FNKAETITAKDGRDLTIGNSDNGSGAEAKTVTFNNVYKSKISADGHNVTNLNSKVKTSSN 1109
Db 985 ANRQDEQRDR-----GSAGVRVVT-----GSDLTVDK-----GEG 1019
QY 1110 GGRESNDNDTGLT-----ITAKNVEVNDITSLKVTNITASEKVTITAGSTINATNGKA 1164
Db 1020 GTORSNSASQAVTGSIDAANGINNVKKD-----ALVGTALNGCRGKT 1064
QY 1165 SITTKGTDI-----SGTISGNTVSATVLDLTKSGSKIEAKSGEANTVTSATGTIG 1215
Db 1065 AVNA-GGDLRLQASDKQSESRSGFNVKASA-----KGGFTADSKNFCA-----GFGG 1111
QY 1216 GTISGNTVNVNAGDLTVYNGAEINATEGAATLTATGNTLTTEAGSSITSTKGOVDLLA 1275
Db 1112 GTHNGESSSTAQVNGNISQOQVELKA-----GRDLTLQ-----TDVKSQGDV-- 1155
QY 1276 QNGSIAGSIANAAN-VTLNITGTLTIVAGSDIKATSGTLVINAKDKLNGD-----ASGDS 1329
Db 1156 -----SUSAGNKVALQNAESTQT-----RKESKLSGNDLGLGSSDS 1192
QY 1330 TEVNAVNASGSGS--VTAATSSSVNITGDLNVTNGLNITISKDGRTVRLRKEIEVKYIQ 1387
Db 1193 KEKTGNSLSAGGAFDIKVNESATERQG-----ATIASDGKVTLSANGKGDALHLQ 1244
QY 1388 PG-VASVEVEIAKR--VLEKVKDLSDERETIAKLGVSAVR-----FVEPN 1431
Db 1245 GAKVSGGSAALEAKNGGILLSEAK---NEOHKDNWSLGIKANAKGGQTFNKDAGGKVDPN 1301
QY 1432 NTITVNT-----QNEFTTRPSSQVIITSEKACFSSGN-----CARV 1467
Db 1302 TGKDPTHTLGLAGLVKGVQEQDKTHANTG--ITAGDVTLSNGKDRLAGARV 1350

```

RESULT 15 OMP_B-RICPR

```

ID OMPB-RICPR STANDARD; PRT; 1643 AA.
AC Q53020; Q9ZC40;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
DE OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN-Breil;
RX MEDLINE-91045972; PubMed-2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RL truncated identical homolog in Rickettsia typhi."
RN Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RP SEQUENCE FROM N.A.
RP STRAIN-Breil;
RA Moron C.G., Yu X.J., Walker D.H.;

```

```

RT "Sequence analysis of ompB of Rickettsia prowazekii."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-Madrid E;
RX MEDLINE-99039499; PubMed-9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria."
RL Nature 396:133-140(1998).
RN [4]
RP PARTIAL SEQUENCE.
RX STRAIN-Breil;
RX MEDLINE-92114896; PubMed-1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RL prowazekii."
RL Mol. Immunol. 29:95-105(1992).
RN [5]
RP CLEAVAGE SITE.
RX MEDLINE-92104668; PubMed-1729180;
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RL deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M37647; AAA26390.1; ALT_INIT.
CC EMBL: AF161079; AAD4234.1;
CC EMBL: AJ235273; CAA15140.1;
CC InterPro: IPR003858; rOmpA_rOmpB.
CC Pfam: PF02708; rOmpA_rOmpB; 1.
CC Antigen: S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 V -> A (IN STRAIN BREIL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREIL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREIL).
FT CONFLICT 178 179 AA -> VC (IN REF. 1).
FT CONFLICT 191 201 TQEAFLPLGA -> INSRSSSYHLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392B6346CC CRC64;

```

Query Match

Best Local Similarity 5.7%; Score 419; DB 1; Length 1643;
Matches 327; Conservative 218; Mismatches 585; Indels 464; Gaps 76;

```

QY 45 LALKP-LSAMLLSLCVTSIPQSVLASGLQGMVVGHTATMQVDGNTKIIRNSVDATINWK 103
Db 1 MAQNFELKKIISAGLVTAATIVAGSGVAM---GAAMQVNRNTNAAATFDGI----- 53
QY 104 QFNIDQ-----NEWQFLOENNNSAVFNVRTSNQISQLKGLDSNGQVFLNPN- 152

```

```

Db 54 --GFDQAAGANIPVAPNSVI--TANANNPTFN--TPN-----GHLNS--LFLEDTAND 98
QY 153 -GITIGKDALINNGFTASTLDISNENIKARNFTFOTKDKALAEIVNGLITVKGDSV 211
Db 99 LAVTINED---TTLGF-----JNTAQAQFFNFVAACKIL-NITGOGITVOEASNTI 148
QY 212 NLIGKVKNGGVISVNG-----GSLLAGOKITSDIINPTITVSIARAPEANVLGD 265
Db 149 NAQNALTKVHGGAANANDLSGLSITFAAPSIVLEFNLPN---TOEAPLTLGANSKI 205
QY 266 IFAKGGNINVRATIR-----NOGKISADSUSVK-----293
Db 206 VNGNGTLLNTNGFIQVSDNTFAGIKTINIDDCOGLMFNSTPDAANTLNIQVGGNTINFN 265
QY 294 --DKSGNIVLSAKBGEA-----EIGGVISAQNOQAKGKLMITGDK-----332
Db 266 GIDGTGLVLVSKNGAATEFNVTGLGNLKGIIEL-NTAAVAGKLIISQGAANAVIGTD 324
QY 333 -----VTLKTGAVIDLSGKEGETYL-----GGDERGEGKNGIOLAKKTSLEKG 376
Db 325 NGAGRAAGFIVSDNGNAATISGOVYAKNMVIOQSANAGQVTEHIVDVGLGCTINF-KT 383
QY 377 STINVSKEKGEA-----IVMGDIALIDGNINAOGSGDIAKTGGFVETSGHDLF 426
Db 384 ADSKVIITENSFGSTNFGNLDTOIVVPDKILKGNF-----IGDVKNNG--NTAGVITF 436
QY 427 IKDRAIVDAKEWLLDFDNVSNIAEDPLENFTGINDPEPTGTGEASDPKKNSELKTLTNT 486
Db 437 NANGALVSA-----STDPNIAVTNINAIIEAEGGVVE-----LSGI 472
QY 487 TISNY-LKNAMTMITASRKUTVNSINSNLSHLLHSKQORGGGVQIDGD--ITSKGG 543
Db 473 HIAELRLGNGSI-----FKLADGTVINGPVNQALMNNALAAAGSIOLOGSALITGIG 527
QY 544 NLTIYSGGVVDVHKNTLDOGLNITASVAFEGGNKARDAANAQIVAQ---GTVTITG 600
Db 528 N-----GGVNAALQHITL-----ANDASKILALDGANIIGANVGA-IFHOANGGTIKLTN 577
QY 601 EGKDFRAN---NVSLNGTGKGLNISSVNNLTHLSGTI-----NI-----638
Db 578 TONNIVWNFDLITTDKTGV-VDASSLTNNQTLTNGSIGTVVANTKTLAQLNIGSSKTI 636
QY 639 --SGNITINOTRKNKTSYQWOTSHDSHWNVSALNLETGANFTFIKYSISNSKG-----688
Db 637 LNAQDVAINELVIENNGSVQLNHT-----YLITKTINAANOQIIVAADP 682
QY 689 LFTQYRSSAGVNFNGVNGNMFNLKEGAKVNFKLKPNENMNTSKPLPIRFLAN---ITAT 745
Db 683 LNTNTLADGTLNLSAENPLS-----TIHFATK-----AANADSILNV 720
QY 746 GGSVFFDIYANHSORGAEKLMSEINISGANFTLNSHVGDGDAFKINKDLTINATNSF 805
Db 721 GKG---VNLIAANNITN-DANVSLHPSRGGTSIVSGTVGGOOGHKLNLLDNGCTTVKF 776
QY 806 -----SLQTKDDFDVGYARNAINSTYNSILGN---VTLGGQNSSSI 847
Db 777 LGDTFTNGGKTEGKSILQISNNYTTDHSVESADN-TGTFEVNTDPTITVILNKOGAFEGV 835
QY 848 -----TGNITIEKAANVTLEANNAPNOQNIRDRVILKGLSLVNGSLSTGENADIKG 899
Db 836 LKQVIISGPGNIVFNEIGNVIVHGAAN-----SISFENASLGT 875
QY 900 NLTISESATFKGTRDTLNI---TGNFT--NNGTAEINITQGV---VKLGNVTND-----946
Db 876 SLFLP-----SGTPLDLVLTIKSTVGNGTVDFNENAPIVVVSIGDSMINNGQIIGKKNIIA 930
QY 947 ---GDLNITTHAKRNORSIIGGDIINKKGLSLNITDSNDAEIQIGGNISOKEG-----N 997
Db 931 LSLGSDNSITVNAVNTLYSGI-----RTTKNNOGTVTLSSGMPNPNPGTIYIGIG 977
QY 998 LTISSDKINIKQITIKKIGID--GEDSSSDATSNANLTIKTKELKLTEDLSISGPNKAEI 1055

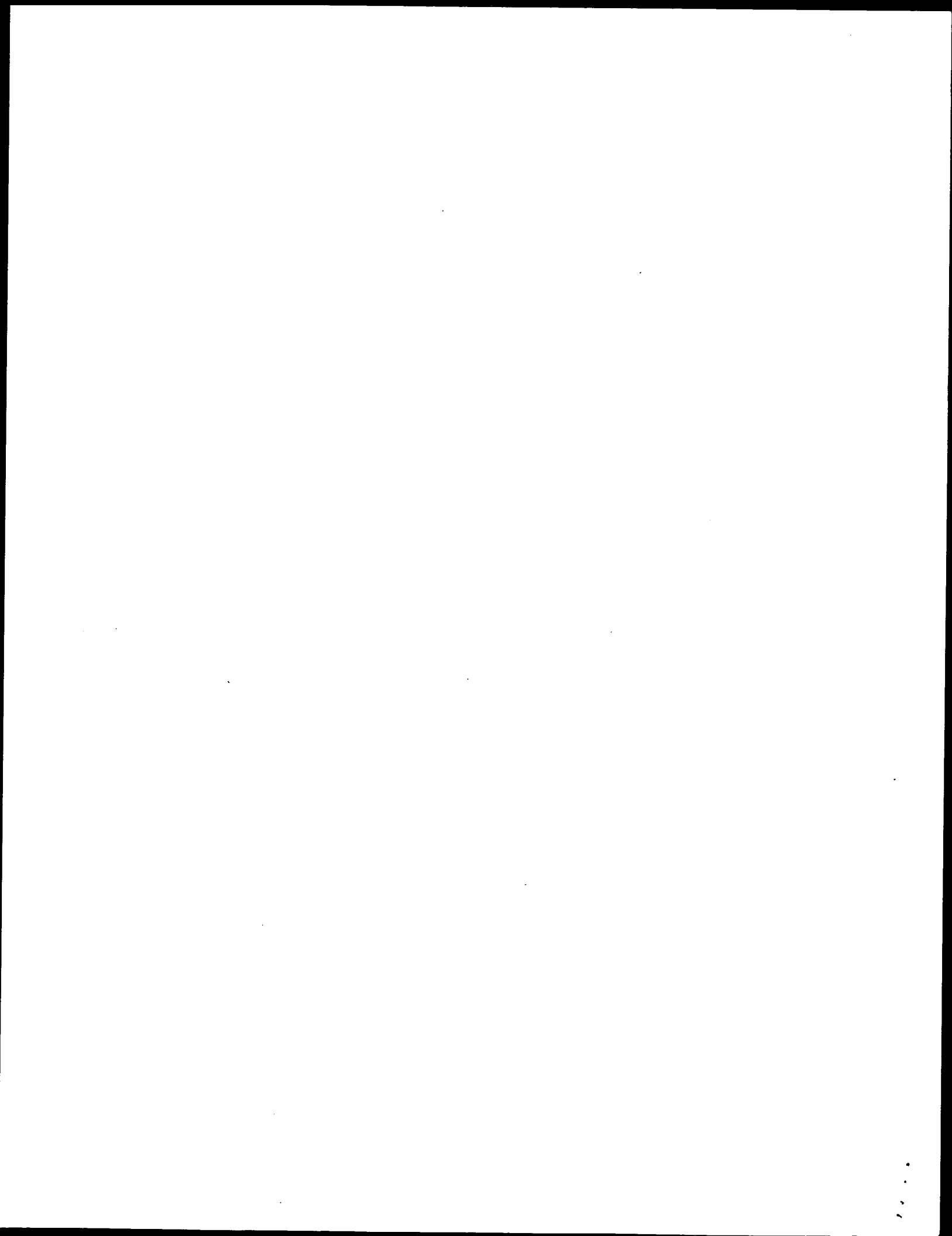
```

```

Db 978 LENGSPKL---KQVFTTTDYNLGSIIANNVTINDYVTLTT-----GGIAGTDF 1023
QY 1056 TAKDGRDLTIGNSDNGSGAEAKTVTFENNVDKSKISADGHNVNLTNSKVKTSSSNGG---- 1111
Db 1024 DAK---ITLGSVN-GNANVREVDSTFSDPRSMIVATOANKGTVTYLGNALVSNIGSLDT 1078
QY 1112 -----RESNDNDTGL--TITAKNVEVNKDIITSLKTVNITASEKVTITTAGSTINATNGKA 1164
Db 1079 PVASVRFTEGNDSGAGLOGNIYSONIDFG-----TYNLTILNSNVILGGGT-TAINGEI 1130
QY 1165 SITTKTGDISGTIS--GNITVSYSATVDLTTKS-GSKIEAKSGEANTVSATGTIGGTISGN 1221
Db 1131 DLTNNLIPANGTSTWGDNTSISTTLNVSSGNIGQVVAEDAQVNAATT-TGTTTIKIQDN 1189
QY 1222 -TVNVTANAGDLTVNGAGAEINATEGANLTATG-----N 1254
Db 1190 ANANFSGTQAYTLIOGGAREFNGTLGAPNFAVTVGSNIFVKYELIRDSNQDYVLTTRTNDVLN 1249
QY 1255 TLTTEAGSITSTKGOVD-----LLAQNGSIAGSINAANVTINTTGT 1296
Db 1250 VVTVAVGNSAIAANAPGVSONISRCLESTNTAAYNNMLLAKDPDVFVGA-IATDTSAA 1308
QY 1297 LTTVAGSDIKATS-----GTL--VINAKDAKLNGDA-----SGDSTEV-----1332
Db 1309 VTTVNLDNTQKTQDILLSNRLGTLRLYLSNAETSDVAGSATCAVSSGDEAEVSYGVNAKPEY 1368
QY 1333 NAVNASGSGSVTAATSSSVNITGDLNTVNGLNII 1366
Db 1369 NIAEQDKKGGIAGYKAKKTGTVGVVGLDTLASDNLM 1402

```

Search completed: March 24, 2003, 15:22:57
Job time : 36 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:22:34 ; Search time 47 Seconds
(without alignments)
6475.140 Million cell updates/sec

Title: US-10-092-880-4
Perfect score: 7407
Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	7318	98.8	1477	2	Q48028 haemophilus
2	4957	66.9	1536	2	Q48031 haemophilus
3	4711	63.6	1557	2	Q9RNI2 haemophilus
4	900	12.2	2055	2	Q93DC7 haemophilus en
5	876	11.8	1910	16	Q82BY3 yersinia pe
6	785	10.6	1371	16	Q8XQ42 raietonia s
7	744.5	10.1	2154	16	Q9HYG6 pseudomonas
8	636.5	8.6	1417	16	Q9HVG6 pseudomonas
9	602	8.1	3705	16	Q8ZHA1 yersinia pe
10	594	8.0	4919	2	Q9F285 yersinia pe
11	593.5	8.0	4919	2	Q9ZHL0 haemophilus
12	561	7.6	2737	16	Q8XP07 pseudomonas
13	553.5	7.5	1018	16	Q9HW06 pseudomonas
14	548	7.4	3930	16	Q9BE20 rhizobium l
15	543	7.3	3165	16	Q8RDQ9 fusobacteri
16	542.5	7.3	2143	16	Q8RIP5 fusobacteri

17	538.5	7.3	1881	16	Q8RGK2	Q8rgk2 fusobacteri
18	531.5	7.2	1270	16	Q8XAN9	Q8xan9 escherichia
19	530	7.2	3501	16	Q8Y106	Q8y106 raietonia s
20	529	7.1	3552	16	Q8XSD6	Q8xsd6 raietonia s
21	523.5	7.1	4152	2	Q9ZHL3	Q9zhl3 haemophilus
22	520	7.0	1975	16	Q9K0S7	Q9k0s7 neisseria m
23	520	7.0	2535	16	Q8ZDR6	Q8zdr6 yersinia pe
24	517.5	7.0	1995	16	Q9JY23	Q9jy23 neisseria m
25	512.5	6.9	2340	16	Q9ZD91	Q9zd91 rickettsia
26	508	6.9	1604	2	Q9KK99	Q9kk99 rickettsia
27	508	6.9	2015	16	Q9JRD2	Q9jrd2 neisseria m
28	505.5	6.8	2712	16	Q9F3X5	Q9f3x5 pasteurella
29	504	6.8	3295	16	Q8Z1A8	Q8z1a8 yersinia pe
30	501.5	6.8	1480	2	Q8VNR0	Q8vmr0 photorhabdu
31	495.5	6.7	3155	16	Q8XYI3	Q8xyi3 raietonia s
32	491	6.6	3233	2	P71401	P71401 haemophilus
33	489	6.6	3241	2	Q9AHF9	Q9ahf9 fusobacteri
34	485	6.5	2026	16	Q9ED50	Q9ed50 xylella fas
35	482.5	6.5	3322	16	Q8XQZ5	Q8xqz5 raietonia s
36	480.5	6.5	2039	16	Q8ZNS7	Q8zns7 salmonella
37	480.5	6.5	1594	2	O32608	O32608 edwardsiell
38	480	6.5	2462	16	Q8RG23	Q8rg23 fusobacteri
39	478	6.5	2751	16	Q8XUK0	Q8xuk0 raietonia s
40	478	6.5	1487	16	Q8YK40	Q8yk40 anabaena sp
41	477.5	6.4	1749	16	P73032	P73032 synechocyst
42	475.5	6.4	1749	16	Q8XZT1	Q8xzt1 escherichia
43	474	6.4	5291	16	Q8XZU1	Q8xzul anabaena sp
44	472.5	6.4	1152	16	Q8YJ30	Q8yj30 neisseria m
45	471	6.4	2514	16	Q8YJ30	Q8yj30 neisseria m

ALIGNMENTS

RESULT 1

Q48028 PRELIMINARY: PRT; 1477 AA.
AC Q48028;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Adhesin.
GN HMW2A
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12;
RX MEDLINE=92192797; PubMed=1548058;
RA Barenkamp S.J., Leininger E.;
RT "Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis."
RT Infect. Immun. 60:1302-1313(1992).
RL EMBL; U08875; AAA20524.1;
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001069; SHprot_acsite.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 1477 AA; 154473 MW; B057C23F1AD24B0E CRC64;

Query Match 98.8%; Score 7318; DB 2; Length 1477;
Best Local Similarity 99.1%; Pred. No. 1.7e-231;
Matches 1464; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGESEKPARMKVRLHAKPLSAMLISLQVT 60
Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGESEKPARMKVRLHAKPLSAMLISLQVT 60
QY 61 SIPOSVLASGLQGMVDVHGTATMQVDGNKTIIRNSVDATINNKQFNIDQENVQFLOENN 120

Db 61 SIPOSVLASGQMDVYVHGATMVDGKNTIIRNSVDALINNKQFNIDQNMVQFLOENN 120
QY 121 NSAVFNRTVSNQISQKGLDLSNGQVFLNPNIGITIGKDAIINTNGFTASTLDISNENIK 180
Db 121 NSAVFNRTVSNQISQKGLDLSNGQVFLNPNIGITIGKDAIINTNGFTASTLDISNENIK 180
QY 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGQKIT 240
Db 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGQKIT 240
QY 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300
Db 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300
QY 301 LSAKEGEABEIGGVIISAQOQAKGGLMITGDKVTLTKTGAVIDLSGKEGGEYILGGDERGE 360
Db 301 LSAKEGEABEIGGVIISAQOQAKGGLMITGDKVTLTKTGAVIDLSGKEGGEYILGGDERGE 360
QY 361 GKNGIOLAKKTSLEKGSINISGKEKGGFAIVWGDIALIDGNINAQSGDIAKTGGFVET 420
Db 361 GKNGIOLAKKTSLEKGSINISGKEKGGFAIVWGDIALIDGNINAQSGDIAKTGGFVET 420
QY 421 SGHDLFIKDNAIVDAKELLDNDVNSINAEDPLENNTGINDFEPTGTGEASDPKKNSELK 480
Db 421 SGHDLFIKDNAIVDAKELLDNDVNSINAEDPLENNTGINDFEPTGTGEASDPKKNSELK 480
QY 481 TLTNTTISNLYKNAWTNITASRKLTVNSNINISGSHLILHSKGQGGVQIDGITS 540
Db 481 TLTNTTISNLYKNAWTNITASRKLTVNSNINISGSHLILHSKGQGGVQIDGITS 540
QY 541 KGGNTIYSGGWDVHKNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQGTVTITG 600
Db 541 KGGNTIYSGGWDVHKNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQGTVTITG 600
QY 601 EKGDFRANVNSLNGTSGKGLNIISSVNNLTHNLSTINISGNITINOTTRKNTSYWQTSHD 660
Db 601 EKGDFRANVNSLNGTSGKGLNIISSVNNLTHNLSTINISGNITINOTTRKNTSYWQTSHD 660
QY 661 SHWVNSALNLETGANFFIKYIISNSKGLTQYRSSAGVNFNGVNGMNSFLKREGAKVNF 720
Db 661 SHWVNSALNLETGANFFIKYIISNSKGLTQYRSSAGVNFNGVNGMNSFLKREGAKVNF 720
QY 721 KLKPNENMTSKPLPIRLANITATGGSVFDDIYAHNSGRGAELKMSINISNGANFTL 780
Db 721 KLKPNENMTSKPLPIRLANITATGGSVFDDIYAHNSGRGAELKMSINISNGANFTL 780
QY 781 NSHYRGDDAFKINKDLINATNSNFSURQTKDDFYDGYARNAINSTYNTSILGNNVTLG 840
Db 781 NSHYRGDDAFKINKDLINATNSNFSURQTKDDFYDGYARNAINSTYNTSILGNNVTLG 840
QY 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIIRDRVTKLGSLLVNGSLSTGENADIKGN 900
Db 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIIRDRVTKLGSLLVNGSLSTGENADIKGN 900
QY 901 LTISEATFGKTRDTLNTGFTNNGTAEINITQGVVKGNGVNTDGDNLNITHAKRNOR 960
Db 901 LTISEATFGKTRDTLNTGFTNNGTAEINITQGVVKGNGVNTDGDNLNITHAKRNOR 960
QY 961 SIIGGDIINKKGSINTDSNNDIAIIGGNIISQEGNLISSDKINITHQITIKKGIDGE 1020
Db 961 SIIGGDIINKKGSINTDSNNDIAIIGGNIISQEGNLISSDKINITHQITIKKGIDGE 1020
QY 1021 DSSSDATSNANLITKTELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGSGAEAKTV 1080
Db 1021 DSSSDATSNANLITKTELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGSGAEAKTV 1080
QY 1081 TFNNVDSKISADGHVNTLSKVKTSNGGREGSNDNDTGLTITAKNVEVNDITSUKT 1140
Db 1081 TFNNVDSKISADGHVNTLSKVKTSNGGREGSNDNDTGLTITAKNVEVNDITSUKT 1140
QY 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTITKSGSKIEA 1200
Db 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTITKSGSKIEA 1200

Db 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTITKSGSKIEA 1200
QY 1201 KSGEANTVSATGTIGTIGTISGNTVNTVNTAGDLTVGNGAEINATGGAATLTATGNTLTTEA 1260
Db 1201 KSGEANTVSATGTIGTIGTISGNTVNTVNTAGDLTVGNGAEINATGGAATLTATGNTLTTEA 1260
QY 1261 GSSITSTKGGVDDLLAQNQSGTAGSINAANVTINTTGTTLTTVAGSDIKATSGTLVINAKDAK 1320
Db 1261 GSSITSTKGGVDDLLAQNQSGTAGSINAANVTINTTGTTLTTVAGSDIKATSGTLVINAKDAK 1320
QY 1321 LNGLDASGDSSTEVNANVNASGSGSVTAATSSSVNITGDLNTVNGLNILSKDGRNTRVLRGKE 1380
Db 1321 LNGLDASGDSSTEVNANVNASGSGSVTAATSSSVNITGDLNTVNGLNILSKDGRNTRVLRGKE 1380
QY 1381 IEVKYIQPGVASVEEVTEAKRVLEKVKDLSDERETLAKLGVSAVRFEVPEPNNTITVNTON 1440
Db 1381 IEVKYIQPGVASVEEVTEAKRVLEKVKDLSDERETLAKLGVSAVRFEVPEPNNTITVNTON 1440
QY 1441 EFTTRPSSQVLIISGKACFSSGNGARVCTNVADGGP 1477
Db 1441 EFTTRPSSQVLIISGKACFSSGNGARVCTNVADGGP 1477
RESULT 2
Q48031 PRELIMINARY; PRT; 1536 AA.
AC Q48031; ID Q48031; PRELIMINARY; PRT; 1536 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Adhesin.
DN HMWIA.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12;
RX MEDLINE=92192797; PubMed=1548058;
RA Barenkamp S.J., Leininger E.;
RT "Cloning, expression, and DNA sequence analysis of genes encoding
RT nontypeable Haemophilus influenzae high-molecular-weight surface-
RT exposed proteins related to filamentous hemagglutinin of Bordetella
RT pertussis.";
RL Infect. Immun. 60:1302-1313(1992).
DR EMBL: U08876; AAA20527.1; -;
DR InterPro: IPR000169; ShpProLacsite.
DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 1536 AA; 159916 MW; 5CAIC31F9DCF188E CRC64;
Query Match 66.9%; Score 4957; DB 2; Length 1536;
Best Local Similarity 67.3%; Pred. No. 2.le-154;
Matches 1046; Conservative 146; Mismatches 285; Indels 98; Gaps 20;
QY 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKSGSEKPAKRVHRLALPLSAMLISLQVT 60
Db 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKSGSEKPAKRVHRLALPLSAMLISLQVT 60
QY 61 SIPOSVLASGQMDVYVHGATMVDGKNTIIRNSVDALINNKQFNIDQNMVQFLOENN 120
Db 61 SIPOSVLASGQMDVYVHGATMVDGKNTIIRNSVDALINNKQFNIDQNMVQFLOENN 120
QY 121 NSAVFNRTVSNQISQKGLDLSNGQVFLNPNIGITIGKDAIINTNGFTASTLDISNENIK 180
Db 121 NSAVFNRTVSNQISQKGLDLSNGQVFLNPNIGITIGKDAIINTNGFTASTLDISNENIK 180
QY 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGQKIT 240
Db 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGQKIT 240
QY 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300
Db 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300

Db 241 ISDIINPTITYSIAAPENAVNLGDIKFAKGNINVRAATIRNOCKLSADSVSKDKSGNTV 300
QY 301 LSAKEGAETGGVTSQAQOAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGERGE 360
Db 301 LSAKEGAETGGVTSQAQOAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGERGE 360
QY 361 GKNCIOLAKTSLKSTINVSCKEKGPAIVMGDIALDGNINAOGSDIAKTGGEVET 420
Db 361 GKNCIOLAKTSLKSTINVSCKEKGPAIVMGDIALDGNINAOGSDIAKTGGEVET 420
QY 421 SGHDLFKDNAIVDAKWLDFDNVSNADPLFNVTGINDPFTGTG-EASDPKNSL 479
Db 421 SGHDLFKDNAIVDAKWLDFDNVSNADPLFNVTGINDPFTGTG-EASDPKNSL 479
QY 480 KTLTNTTISNYLKNWMTNITASRLKLTVNSNINISNSHLIHSKQGGVQIDGDI 539
Db 479 KTLTNTTLESILKKGFTVNITANORIYVNSNL-SNGSLTLWSEGRSGGVIEINDIT 537
QY 540 ----SKGNLTISGGVWDVHKNTLD-QGFLNITA-ASVAFEGGNKARDAANAKIVAQ 593
Db 538 TGDTRGANLTISGGVWDVHKNTISLGAQGNINITAKQDIAFEKGSQV-----ITGQ 590
QY 594 GTVTITGEGKDFRANNVSLAGTKGLNIISVNN---LTHNLSTINISGNITINOTTRK 650
Db 591 GTIT-SGNQKGFENNYSLAGTSGLOFTTKRTNKYAITNKFECTNLISGKVNISMVLPK 649
QY 651 NTS-YWQTSDDSHWNVSALNLETGANFTF-IKYISSNSKGLTQYRSSAGVNFNGVNGM 708
Db 650 NESGYDFPKGRTYWNLTSLNVSSEGEFNLATIDSRGSDSAGTLTOPYNLNGISF---NKDT 706
QY 709 SFNLKEGAKVNFELKPNENMTSPLP-IREPLANITATGGSGVFEDYIYANHSQ---RGAE 764
Db 707 TFNVARNARVFDIKAPIGINKYSLNVSFNGNISVSGGSVDFTLLASSNNVOTPGV 766
QY 765 LKMEINISNGANTLNSHVGRDFAKINKDLTINATNSFSLRQTKDDFVGVARNAIN 824
Db 767 INSKYFNVSTGSSLRFTSGTKTGFSTIEKDLTLNATGGNTLLQVEGT--DGMIGKGV 824
QY 825 STYNISILGGNVTLGGONSSITGNITIEKAAVNTLEANNAPNOQIRORVILKGLLV 884
Db 825 AKKNTFEGGNITGSKRAVTEIEGNVTINNNANVTLIGSDFDNQ--KPLTIKKVLIIN 882
QY 885 NGSLTGENADIKGNLTISSEATFKGTRTLNITGNFTNNGTAENITGVGVKLVNT 944
Db 883 SGNLTAGGNIIVNAGLTVESNAFKAITNFFNVGGFLDNKGNISISIAKGGARFKDID 942
QY 945 NQDGLNITHAKRNORSIIIGDINKKGLSLNITSDNDAETIOIGNISQKEGNLTISDDK 1004
Db 943 NSKNLSITNSSTYRTIISGNITNKGDLNITNEGSDTEMQIGDVSQKEGNLTISDDK 1002
QY 1005 INITKOITIKKIDGEDSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLT 1064
Db 1003 INITKOITIKAGVDGENSDATSNANLTIKTKELKLTODLNSGFNKAETAKDGRDLT 1062
QY 1065 IGNSNDGSGABAKTIVTFNNVNDKSKISADGHNVTLSNKKVKTSSSNGGRESNDNDTGLTI 1124
Db 1063 IGTNTSAD--GTNAKVTFTNOVKDKSKISADGKHVTLHSKVFSGSNNTEDSDNNAGLTI 1121
QY 1125 TAKNVENKDIITSLKTVNITA--SEKVTITAGSTINATNGKASITTKT----- 1170
Db 1122 DAKNVTNNNITSHKAVSISATSGEITTKGTITNATTGNVEITAQTSILGGIESSGS 1181
QY 1171 ----- 1171
Db 1182 VTLTATEGALAVNSIGNVTVTANSALTTLAGSTIKGTESVTTSSQSGDGGTISGCT 1241
QY 1182 VSVSATVDTLTKSGSKIEAKSGEANTVSATGTIGCTISGNTVNTANAGDLTVNGAEIN 1241
Db 1242 VEKATESLITGOSKIKATGEANVTSAITGIGTISGNTVNTANAGDLTVNGAEIN 1301
QY 1242 ATEGAATLTFATGNTLTTEAGSSITSTKQGVDDLLAQNQSGTAGSINANVTNLTGTLTVA 1301
Db 1302 ATEGAATLTSKGLTTEASSHITSKAGOVNLSAQDQSVAGSINAANVTNLTGTLTIVK 1361

QY 1302 GSDIKATSGTLVINAKDAKLANGASDSTEVENAVNSAGSGSVTAATSSSVNITDGLNTVN 1361
Db 1362 GSNINATSGTLVINAKDAELNGHTVNVNATNANGSGSVIATSSRVNITDGLIN 1421
QY 1362 GLNIIISKGRNTVRLRGKEIEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLG 1421
Db 1422 GLNIIISKGINTVLLKGVKIDVKYIQPGIASVDEVEIAKRVLEKVKDLSDEERETLAKLG 1481
QY 1422 VSAVRFPENNTITVNTQNEFTTRPSOVIISEGKACFSSGNGARVCTNVADGQ 1476
Db 1482 VSAVRFIEPNNTITVDTQNEFATRLSRIVISEGRACFSNSDGTATVCVNIADNGR 1536
RESULT 3
Q9RNI2 PRELIMINARY; PRT: 1557 AA.
AC Q9RNI2
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
DE HmWA.
GN HmWA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A950006;
RX MEDLINE=20359363; PubMed=10899870;
RA van Schilfgeaarde M., van Ulsen P., Eijk P., Brand M., Stam M.,
Kouame J., van Alphen L., Bankert J.;
RT "Characterization of adherence of nontypeable Haemophilus influenzae
to human epithelial cells.";
RL Infect. Immun. 68:4658-4665(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A950006;
RA van Schilfgeaarde M., Eijk P., van Ulsen P.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF180944; AAD56660.1; --
DR InterPro: IPR000169; SHprot.acsite.
DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 1557 AA; 160461 MW; 62524D384398CB37 CRC64;
Query Match 63.6%; Score 4711; DB 2; Length 1557;
Best Local Similarity 63.3%; Pred. No. 2 3e-146;
Matches 1003; Conservative 155; Mismatches 290; Indels 136; Gaps 23;
QY 1 MNKIYRLKFKSRNALYAVSELARGCDHSTKESKPEARMKVRHLAKPLSAMLISLGV 60
Db 1 MNKIYRLKFKSRNALYAVSELARGCDHSTKESKPEARMKVRHLAKPLSAMLISLGV 60
QY 61 SIPOSVLASGLQGMVHVGTATMVDGNKTLIRNSVDALINWKOFNIDQEMVQFOENN 120
Db 61 SIPOSVLASGLQGMVHVGTATMVDGNKTLIRNSVDALINWKOFNIDQEMVQFOENN 120
QY 121 NSAVFNRTVNSQISQLKGLDSNGOVFLINPNTIGTKDALINTNGTASTLDISNENIK 180
Db 121 NSAVFNRTVNSQISQLKGLDSNGOVFLINPNTIGTKDALINTNGTASTLDISNENIK 180
QY 181 ARNFTFQTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGVIISVNGGSSISLAGOKIT 240
Db 181 ARNFTFQTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGVIISVNGGSSISLAGOKIT 240
QY 241 ISDIINPTITYSIAAPENAVNLGDIKFAKGNINVRAATIRNOCKLSADSVSKDKSGNIV 300
Db 241 ISDIINPTITYSIAAPENAVNLGDIKFAKGNINVRAATIRNOCKLSADSVSKDKSGNIV 300
QY 301 LSAKEGAETGGVTSQAQOAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGERGE 360
Db 301 LSAKEGAETGGVTSQAQOAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGERGE 360

QY 361 GKNGIOLAKTSLEKSTINVSKEKGGFAIWMGDIALIDGNIQAQSGDIATGGFVET 420
Db 361 GKNGIOLAKTSLEKSTINVSKEKGGFAIWMGDIALIDGNIQAQSGDIATGGFVET 419
QY 421 SGHDLFIKDAIYDAKEWLLDFONVINAEDPLFNWNTGDEPTGTGASDPK-NSEL 479
Db 420 SGHLSIGDAVAEAREWLLDPONVIT-----SNGNDQSQLKDDRGDSFNKILADN 471
QY 480 KTLTNTTISYLNKNAWMTNITASRKLTYNSSINISHLIHSKQGGGVQIDGDI 539
Db 472 KHTVNNKTLSTALAKGIVNISAKKKVNTADIN-VHGTLLHSE---QGGVEINGDIT 527
QY 540 S-KGGNLTISYSGWVDVHKNITLDQGLNITA-ASVAFE-GGNKARDAANAKIVAQGT 596
Db 528 SEONGMLTITKAGSWVDVHKNITIGTFLNITAGGSVAFEKAGDGKRAASDAKIVAQ 587
QY 597 TITGEGKDFRANNSUNGTKGLNIIS---SYNNLTHNSGTINSGNITINOTTRKNTS 653
Db 588 T-AGSQDFRNNVSLNGTGRGLKFTAKGNKGNFSKAFDGVNLTSGNISINHTANNQLS 646
QY 654 YWQTHSDSHWVNSALNLETFANF--TFIK-----YISSNSKGLTQYRSSAGVNF 704
Db 647 YFHOGYTYWNLTLQNVSDSSFSLSITKDAIKVGYDNAKDK-----KNTGGIGF--- 697
QY 705 NGNMFNLKEGAKVFKLK-PNENMNTSKPLRFLANITATGGSV---PFDIYANHS 760
Db 698 TRDTTFNKGARVDISYTLPLSPVKNRSIAAANVDGNITVKGGVNLFKNALSNNYKT 757
QY 761 RGAELKMSSEINSGANFTNLNHSVGRDGAFAKINKDLTINATNSFSLRQTKDFDVGAR 820
Db 758 PCVNISRFINVTESQNLNITGSMFTLNFVANDLIINATNSFVSIKEIEGT--DTHLD 815
QY 821 NAINSTYNISILGNNVTLGGQNSSSTGNTITIEKAANVTLEANNAPNOQNIRDRV 880
Db 816 TGLKVGNNVTIKGGNVTLGSKNAKTKFKDNKNTVEKGANLTASANFGNH---KGALT 872
QY 881 SLLVNGSLTGENADIKGNITISSEATFKGKTRODLNLTNGFTNNGFAEINITGV 940
Db 873 NINTGKLVATGDTIDVSDFTVGNDAFTNGNNTNNLNTNGFTNNGFTIIDVKGAAKL 932
QY 941 GNVYNDGDLNITTHAKNRQSIIGDIIKNGSLNITSDNDAEIQIGNITSQEGNLT 1000
Db 933 GNITNEGSLNITTHANTNOKIITGNITNKKGLNIRDNKNAEIQIGNITSQEGNLT 992
QY 1001 SSDKINIKQITIKGIDGSSDSDATSNANLTIKTKELKLTEDLSISGFNAKAITAD 1060
Db 993 SSDKVNITKQITIKAGVNGSDSGTENANLTIKTKLETLNLTNINISGPFHAKAITAD 1052
QY 1061 RDLTIGNSDNGSGAEAKVTTFNNVKDSKISADGHNVTLNSKVKTSSNGGREGNSD-ND 1119
Db 1053 SLDIIGKASSDSNAGAKGVIFDKVDSKISAGNHNVTLNSEVTSNGSNAAGDSNGN 1112
QY 1120 TGLTITAKNVEVKDITSLKTVNITASE-KVTTAGSTINATNGKASITTKT----- 1170
Db 1113 AGLTISAKDVANNITSHKTNISATGNVTTKETGTINATGGVEVTAKTGDKGIE 1172
QY 1171 -----GDISGT 1176
Db 1173 SKSGGVTLTATGDTLAVGNISGNTVSVTANSGLTITTKADSTIKGTGVTLLSQSGDIG 1232
QY 1177 ISGNTVSVATVD--LTTKSGSKIEAKSGEANVT-----SATG 1212
Db 1233 ISGNTVSVATVD--LTTKSGSKIEAKSGEANVT-----SATG 1212
QY 1213 TIGGTISGNTVSVATVD--LTTKSGSKIEAKSGEANVT-----SATG 1212
Db 1293 DISGTISGNTVSVATVD--LTTKSGSKIEAKSGEANVT-----SATG 1352
QY 1273 LLAONGSIAGSNAANVTNLTGTLTVAGSDIKATGTLVINAKDAKLDGASGSDTEV 1332
Db 1353 LSARDGNIGGSNAANVTNLTGTLTVAGSDIKATGTLVINAKDAKLDGASGSDTEV 1412

QY 1333 NAVNASGSGVTAATSSSVNITGDLNTVNGLIISKDRNTVLRGKEIEVKYIQPGVAS 1392
Db 1413 NATNASGSGVTAATSSSVNITGDLNTVNGLIISKDRNTVLRGKEIEVKYIQPGVAS 1472
QY 1393 VEEVIEAKRVLEKVKDLSDEERETLAKLGYSVAVRFEVNNITVNTQNEFTTRPSSVII 1452
Db 1473 AEEVIEAKRALEKVKDLSDEERETLAKLGYSVAVRFEVNNITVNTQNEFTTRPSSVII 1532
QY 1453 SEGKACFSSGNGARVCTNNAADDQ 1476
Db 1533 SEGKACFSSGNGARVCTNNAADDQ 1556
RESULT 4
Q93DC7 PRELIMINARY; PRT: 2065 AA.
AC Q93DC7; Q93DC7; Q93DC7; Q93DC7; Q93DC7; Q93DC7; Q93DC7; Q93DC7; Q93DC7; Q93DC7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE RSCA.
GN RSCA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21437624; PubMed=11553561;
RA Nelson K.M., Young G.M., Miller V.L.;
RT "Identification of a Locus Involved in Systemic Dissemination of
Yersinia enterocolitica";
RL Infect. Immun. 69:6201-6208(2001).
DR EMBL; AF394927; AAK77860.1; -
DR InterPro: IPR001950; TIF SU11.
DR PROSITE: PS01118; SU11.1; UNKNOWN.1.
SQ SEQUENCE 2065 AA; 210024 MW; A306D0763682D176 CRC64;

Query Match 12.2%; Score 900; DB 2; Length 2065;
Best Local Similarity 25.0%; Pred. No. 7.8e-22;
Matches 420; Conservative 258; Mismatches 623; Indels 378; Gaps 81;

QY 2 NKIYRLKSKRLNALVAVSELARGC-----DHTEKSEKPKMKVHRLALK 48
Db 3 SKLYKLIIFCRRUGCLIAVGEFTTYGRSFSFGKKIINDNHTRAG-----KLHLAT- 54
QY 49 PLSAMLLSLGVTSPQSVLAS---GLOGMDVVGHTATMOVDGK-TIIRNSVDALINNKQ 104
Db 55 -----LTCLALGTLPLLVFAHPLPVNG-NIVVGOGKMDVNTTITITQSDKLAINWGS 108
QY 105 FNIDONEMVQFLOENNSAVFNRTVSNQISQLKGLDSNGQVFLNPNITIGTKDAINT 164
Db 109 FDIAOGNNVIYVQPGQSIALNQVLGRDASQIYGNLKANGQVFLNPNIGLFGKAQVDV 168
QY 165 NGFTASTLIDISNENIKARNFTFEOTKDKALAEIVNHGLITVKGDSVNLIGGKVKNE--G 222
Db 169 GGLIATSKMSNQDFISGCYTL--TSQKQGLVQANLRTTAGYIALIGQVQDNQPSG 226
QY 223 VISVNGSISLQAGKITIS---DIINPTITYSIAAPENEAVALN-----GDIFAKGNI 273
Db 227 VINTPQKVALASGSRVILNDRNLLGVQV-----QGEQVNTLLQNGGLIRADEGVI 279
QY 274 NVRA-----ATIRNOGKLSADSVSKDGSNTVLS-AKEGEAEIGVISAQNOQAKGG 324
Db 280 QLTAAQKEMLMNTVIDNTIGILQARGLS-EKNGVIYLVNGNEGVSQQGMVNSQQGRGG 338
QY 325 KLMITGDKVTLTKTGAVIDLSGKE-GGETYLGDSBERGEGKNGIOLAKKTSLEKGTINVS 383
Db 339 NVILLIGENIHLVAASKIDARDEGGKVLVGGDWQGGKMKL-IKNARSVMMDKANIDVSS 397
QY 384 KEK--GGFAIWMGD-IALIDGNIQAQSGDIATGGFVETSHDLF-----IKONAI-V-D 434
Db 398 THOGAGGTAVLWSEHYTGfyGDIHARG-GSLSGGGGQVETSSQRNLQSGFRVDSAINGN 456

QY 435 AKWLLDFONVSI---NAEDPLFNNTGINDERPTG---TGEASDPKKKSELKTLTINTTI 488
Db 457 TGRWLLDPAEYVINGSGAESGVSVQIG---DIPAGVVKNAQVFTPMAN---VTQILNTSI 510
QY 489 SNYLNKAWNTNITASR---KLTWNSSIN---IGSNHILHLSKQGGVQIDGD 537
Db 511 NAQDKGTNTVITITSSNLSLNCRCWNTITLQADINKTAGGATLTL---QADGN 560
QY 538 ITSCKGMLTYISGGVDVHKNTILDOG-----FLNITASVAFEGC-----NNKA 582
Db 561 IVSNHNTATG---KUNLNLGSDSIVDSIITLNNSDVLLNGDGLLKHANENNA 615
QY 583 RDAANAKIVAGQVITITGEGKDFRANKVSLNG-TGK-----GLNISSVNNLHNLSTIN 637
Db 616 R-----ISIMGROYGVNLTLEGNTGMAVGVNLSNSAN---ISVAGESTR 658
QY 638 ISG-NITINGQ-----TTRKNTSYWOTSHD-SHNWVSALN---LETGANFTFI 679
Db 659 ISGESSANQGWRGIDISGDSVFAGKGNMSFTMTNSRSSWMTFTNATAGDKNITF- 717
QY 680 KYISSKGLTTOYRSSAGYNF-NG-----VNGNMSFNK-----EGAKVNFK 721
Db 718 ---QANANGST-----SGGVDFNTGSLVSKSGNISDFINGEIIITOTSFGRLIOGSKVSGN 769
QY 722 LKPNENMTSKPLPIREL---ANITATGGGVDFDIYANHSGRGAELKMS---EINISNG 775
Db 770 ---NVNVEINTKGVDFELLRDSHITATAG-----NISANATTTHKGLWISGDDTLNASKN 821
QY 776 ANF---TLNSHVRGDDAFKI---NKDLTIN-ATNSNFSLRQTKDDFDYDGYARNAINSTY- 827
Db 822 IKLOGVTNSTTVGADAIKISGSSSVOVNMAAGNISMAVNVKGTVEG---STISADYT 878
QY 828 NISILGNNVTIGGONSSSITGNITIKKANVTLEANNAPQOINRDIRVTKLSLLVNGS 887
Db 879 NIKAGQDFNLISGVKSGPPSNVNI-SADNIIILNGNISDN---DAVV-----WNTF 927
QY 888 LSLGENADIKNLTIS---ESATFKG---KTRDTLNTGNTNNGTAENINIGGVVKLG 941
Db 928 LAAG---DIKANLSSPNYKALYFRNGGSGAGNLVAVNTSGASVEANTITGIANRM 984
QY 942 NVTNDGDLNITHAKRNORSI-IGDIIIN---KKGSLNITDSNDAEIQICNISQKEG 997
Db 985 NVTVGRDTSIIANNHGNMTGAGIGVDVYVNFPEAKNG--NFTANN-----GSKSIGIAN 1035
QY 998 LTISDKINITKQITIKKI---DGEDSSDATSNANLTIKTKEKLTEDLSISGFNAE 1054
Db 1036 ADIRANEVNLISNTSRADGVVIREANITITLTGNINANATSNRGIVIRQNTLSAQKELM 1095
QY 1055 ITAKDGRDLITGNSNDGSGAEAKTVTFNNYKD---SKISADGHNVTLSKVKTSSSNGG 1111
Db 1096 LAA-----TSSAASEAIIVOGLSDSRSHLVAQG-----NISLKGNSKGS 1136
QY 1112 RESNSDN---DTGLTITAKNVEVNDKITSILKTVNITASEKVTITAGTINATNGKASITTK 1169
Db 1137 NPPRSSVNLNLSVTSVGNKIDIN-----SSS 1162
QY 1170 TGDISTGNTVSVSAVDTLTKSGSKIEAKSGEANVTSATG-----TIGGTISGNTV 1223
Db 1163 VGD---GDYFNVDNLNAGLVNVTYGEALSALSTATNSVLISGNNISIKALNGLICKAI 1220
QY 1224 NVTANAGDLTVNGAEINATEGAATLTATNTLTTEAGSSI-----TSKGGVDLIAQ 1276
Db 1221 NTSQAGAGTLFRANGS-LSVAGNIALTOGTTGATRGIAFYGANTLNIAKDSQLSLGE 1279
QY 1277 N-----GSTAGSINAANVTFLNTGTTL-----TTVAGSDIKATSTVLINAK-DAKL 1321
Db 1280 NTGSOITAGGNGISYLSPLKTLINNNGLTMBEGRSTGAGINFTPGNNTVVLNGBGDSLI 1339
QY 1322 NGDA-SGDSSTEVNAVNASGVSVAATSSS-----VNITCDLNTVNLGN 1364
Db 1340 KGSVAGSGVAISGMVNNSSGPTIEGSSDTGSGVHLFSABHQINRINVTGSSIOAEGLR 1399

QY 1365 IISKDRNTVRLRKEIEVKYIQFCVASVEEVIEAKRVLEKVKDLSDEERTFLAKLGVSA 1424
Db 1400 ISGNATITDALSCKSINGSGIK-----VDSLLSGSVITHVVDLNAALNGSHTGIGVEI 1454
QY 1425 VRFVSP-----NNT-----ITVNTQNEFTTRPSSVOIISSEKACFSSSGARVCTN 1470
Db 1455 TSDINGIHQSIINGTTDGIYGDIDKKNLHVHTGTSETDILLTQGVATTGSGTGKLDGN 1513
RESULT 5
QY 082BY3 PRELIMINARY: PRT: 1910 AA.
AC 082BY3:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE putative adhesin.
GN HMMR OR YFO3247.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE-21470413; PubMed-11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome analysis of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL; AJ414156; CAC92482.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1910 AA: 195076 MW: 665127A098E4045 CRC64;
Query Match 11.8%; Score 876; DB 16; Length 1910;
Best Local Similarity 25.5%; Pred. No. 4.3e-21;
Matches 421; Conservative 235; Mismatches 665; Indels 332; Gaps 79;
QY 2 NKTYRLKSKRLNALVAVSELRGCDHSTKSGSEKPARMKVRHLALKPLSAMLISLGVTS 61
Db 3 SKLYKLIFCRCLLAVGFEFTRSYGRFSSKGGAGNORRAVGILSLRAMTGLALGI 62
QY 62 IPOSVLASGLQMD--VHGTATWQVGNKTIIRNSVDAT-INNKOFNIDONEMVQFLOE 118
Db 63 FPLLVLAHPVLPVNGHVHVIQGGMLDQSSLTVTQOTDKLAINWDSFDTAHGSHVIAQP 122
QY 119 NNSAVFNRTVSNQISOLKILDSNGOVFLINPNTIGTKDALINTNGFTASTLIDISNEN 178
Db 123 GSQSIALNOVGOSASQIYGRQLANGOVFLNPRGILFGKEAQNNGVGLVASTKYMSNPE 182
QY 179 IKARNEFTFEOTKALAEIYNHGLITVKGDSVNLIGGKVKNE--GVISVNGSGISLLAG 236
Db 183 FLSDGYRL--IGGESEGNIIINQANLRSAPGGYIALVGNRIDNQRSGSITTPGQNTVLAVG 240
QY 237 QKITIS----DIINPTITYSIAAPENAV-----NLGDFIPAKGNINVRA-----AT 279
Db 241 HSVTLNLDHGNLLGVQI-----QGETVAALIQNGGLIQADGGVQLTAKGKMLMDTV 293
QY 280 IRNOGKLSADSVSKDSKSNITVL-SAKEGEAEIGSVISAQNOQAKGKMLITGDKVTLKTG 338
Db 294 IDNTGILQAKGLSA-KNGAIIYLDGGEGVVVSQMGITIDVNNQOGRGAVVEGKRIYLNKN 352
QY 339 AVIDLSKEGETYL-CGDERGECKNGIOAKKTSLEKSTINVSKEK--GGFAIVWG-394
Db 353 SNIAKAKTAGGTVLVGGGQWQK--DNQIRNATAVMDKGSIDIVSASRNGPGGSVAVLWSE 411
QY 395 DIALIDGNINAAQ---SGDIATKGTGFVETSGH---DLFIKDNAIV---DAKEMLLEDNV 445

Db 412 DYTGFHGNIRARGGPGSGD-----GGQVETSSQRNIQAQGVQDASAVRGSAGYWLDDPAEV 467
QY 446 SI---NAEDPLFNNTG- INDEPFTGTGEADPKKNSLKTTLTNTTISLYLKNAMTMIT 501
Db 468 TVSSGAESGVMTKVGNIAPAEF-FSSAHIFIPANI---TQILNSINTQLNSGNTVIT 523
QY 502 ASRKLIVNSINIGNSHLILSHKQGGGVQIDGDIYSKGG---NLTIYSGGWVDVHKH 558
Db 524 TS-----NSSLT-----GCQMCNITVQADITKTAGADATLTLQADGNIVWNN 566
QY 559 ITLDQGLNITAASVAFEGGNKARDAA---NAKIVAOCTVITTEGEGKDFRANVNSLNGT 615
Db 567 ITADAGKLNLLA-----GNITADSATLNNKVL-----NGDFLAKHANDNT 613
QY 616 GGLNLIIS---SVNNLTHNLSTINISGNITINOTTRKNTSYWQTSKSHWNVSALNLET 672
Db 614 AR-IGLLGRYDVGNFT---LDGNTALASQGVGNISNAANISV----- 652
QY 673 GANFTFIKIVISSNKLITQYRSSAGVAFN-----GVNGNMFNLKEGAKVNFKLKPN 726
Db 653 -AGEIVISGVSSNRG-----QGWKGIDISNNILITGV-GNMTFSIGNSNVSW-MGAFT 704
QY 727 NNTSKPLPIRELANITATGG-----GSVFEDI-----YANHSRG 762
Db 705 NATITSDKNILFQGTSSGGGVDFVNSRILSKSGRVLFEDINGNIWVKNVGLRVNSOLS 764
QY 763 AELKMEINISGANFTL-NSHV---RGD---DAFKINKDLTINA-TNSNFSLR-----Q 809
Db 765 ADKVFANVTGVDFLLRDSHTVATSGDINANANTINKGIWISGKTNLASGNVNLHV 824
QY 810 TKDDFYDYARNAINSTYINISITLGNVTLGGONSSSITG-----LMTNTELTAKG---DIKTD 934
Db 825 TTNSAYAG--ADAIKISSSSNNVNITAGHISLIAVNGKEIGSVSDVDYANIAKNG 882
QY 850 ----NITTEKAA---NVYLEANNAPNOONI---RDRVIKLGLSLLVNGSLTGENADIKGN 900
Db 883 DFNLNITGKSGFPNNATITANNISMGNITANDAV-----LMTNTELTAKG---DIKTD 934
QY 901 LTI-SESATFKGK---TRDPLNITNGFTNGTAEINITGVVKGVLGNVNDGDLNITHA 955
Db 935 LTPSKGLWFRNGMGWTAANNILLVANSTSGE-----TVKI-NASSSNKNMITA-- 983
QY 956 KRNORSIIGDIIKNGK-SLUNITDSN---NDAEIQIGGNISQKEGNLTISDDKINIKOI 1011
Db 984 -GKDISIAGNSKATGPNINIENVIETNNGFTNGITSTWLSGVNVSANGVDITSNS 1042
QY 1012 TIKKGIDGEDSSDATSNANLTIKTELK-LTEDLSISGFNKAETAKDGRDLT-IGNS 1068
Db 1043 TGTGGI-----VLONTNILLTVGDINTIVTNSCKGIWIKNSNPLNSNKDITLVGVS 1094
QY 1069 NDGSGAEAKTVTFNNYKD---SKISADGHNVTLNSKVKTSSNGGREGNSDN--DTGLT 1123
Db 1095 AGQNEG-----VLIQSSDASRNINISAQG-NITLIGKM-----GNGSCOHSLNLGNVSLT 1144
QY 1124 ITAKNVEVNDKITSKLVNITASEKVTTTAGSTINATNGKASITTKTGDISGTSIGNTVS 1183
Db 1145 SSGRNDINGSAGTGVDVFTNVE-----LNATAGNVSIYAEI-----NTALS 1187
QY 1184 VSATVLDLTKSGKIEAKSEANVTSAETGTTIGTISGNTVNTANAGDLTVNGCAINAT 1243
Db 1188 TSLNAVLSLGNNSIKAQNG-----WLIGKAFNTQAGIGFRANSSLSVDGNILKGET 1242
QY 1244 EGAATLFTAG-----NTLTTEAGSSIT---STKGQVDLLAQAQSTAGSINAAVNTLNT 1294
Db 1243 EGVGA-TRKGIDFYGANTLNIKGSQSLNGENKGAQDTAGGNGISYTSI--AKLIVNN 1299
QY 1295 GTL-----TTVAGSDIKATSGTLVINAK--DAKLNGDA-SGDSTEYNAVNSAGSGSVTAA 1346
Db 1300 GSKMEGRSTSGTGINFPSPNNLTFLVNGDGTLLIKGSSVAGTGAASIVVNNSTGPMTE 1359
QY 1347 TSSS-----VNITGDLNTVNGLNIISKDRNTVRLRGKIEVKYIQ--- 1387
Db 1360 GISTDAGVHLFSAEHRIDRINTVSGSTHAEGLRISGNAIIVDTTLTGKISNGSGVKIDS 1419

QY 1388 -PGVASVEEVTEAKRVLEK--VKDLSDEERTLAKLGVSAVRFVEPNNTIT-----VN 1437
Db 1420 LPGSSV-----TRSLVDNATLNGSSSGKGVETISDINGIHHSINGTTGTGCGYDIDG 1474
QY 1438 TQNEFTTRPSQVITISECKACFSSGNGARVCTN 1470
Db 1475 ENSNVGTGSEADLLILQGVATTGTGTGIKNGN 1507
RESULT 6
Q8XQ42
ID Q8XQ42 PRELIMINARY; PRT; 1371 AA.
AC Q8XQ42;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative hemagglutinin-related protein.
GN RSP1444 OR RS03099.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
Arlat M., Billault A., Bottier P., Camus J.C., Cattolico L.,
Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
Weissenbach J., Boucher C.A.,
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646084; CAD18595.1; -
DR InterPro; IPR002064; DNA_pol_B.
DR PROSITE; PS00116; DNA_POLYMERASE_B; UNKNOWN_1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 1371 AA; 132558 MW; 1D94A3037BF8CF8 CRC64;

Query Match 10.6%; Score 785; DB 16; Length 1371;
Best Local Similarity 24.2%; Pred. No. 2.8e-18;
Matches 365; Conservative 196; Mismatches 533; Indels 414; Gaps 61;
QY 42 VRH-----LALKPLSAMLISLGVTS---IPQSVLASLQGM----- 74.
Db 1 MRHFTLHVALIPSGAAITRAAVTGHARHPRAPRNALGGTVLLGTAAAMPAAAPAG 60
QY 75 -DVVHGATMQVDG-NKTIIRNSVDAILNWKOFNDQEMVQFLQENNNNSAVFNRTSNQ 132
Db 61 GTVTSGSGSIGONGTITITOTSSRLAIDWSAFQAGETVNFQPGAGATALNRVTGHE 120
QY 133 ISOLGILDSNGQVELINPNGITICKDALIINTNGFTASTLDISNENIKARFTFQTKDK 192
Db 121 ATSLGSLNANGTVFLLPNPVGFGAGVNVGVGLVASTLGLSNADFEAGRYAL---SGG 177
QY 193 ALAEIVNHLITVKGDSVNLIGGKVKNEGVSISVNGSGISLLAGOKITISDIINPTIYS 252
Db 178 STAGVTNQTITVPSGGKVALIANSDVGNAGNISAPRGVLAGAGNVTLTLDGSPLYGT 237
QY 253 IAAPNEAV--NLGDIKFAKGN-----NVRAAIT-RNQG--KLSA 288
Db 238 ISAGAARTLVNNGGMIVADGGRVLTARGLDLSSESVNNTTGVVARTVGNNGQTIELIG 297
QY 289 DSVS--KDKSGNVLISAKAGEAEIGGVISAQNOQAKGKLTITGDKVTLTKTCAVIDLSGK 346
Db 298 DPVAGLTQVSGQIDASAPDG-----GG-----NGGSVKVLGTVGVFGAQIDVSGM 344
QY 347 EGGETYL--GGDERGEGKNGIQIAKKTSLSEKSTINVS--REKGGFAIVWG--DIALIDGN 402
Db 345 AGGCTALLIGGNAQAGPE--PNATATYVAPTAADASAIRQNGNGRLIVGWTDVANVHGS 402

Db 518 ISAPILFRNGSLAIRAGGNINFLSGTPTQSGTIVDLSGGLTMMQSTAGKISQOAGT--A 575
QY 583 RDAAN-----AKIVAQGVTTITGEGKDFRANVSLNG-----T 615
Db 576 LIAANLAGRAGSIDLASWMDYAGNALQTFNGTLKVRQSNATGVTSIGVFDFINQSWT 635
QY 616 GKGLNISSVNN--LTHNLSGTINISGNITINQTTKNTSY-----W 655
Db 636 GTAQNVSSVGTIRLEANSVGT--TGNYTL--TADGNSEFDRLVFTALPYRRVSGSASF 690
QY 656 QTSHDHSHVNSALNLE--TGANFTEFK-----YSSNSKGLTT---QYRSSAGV-NFN 702
Db 691 PTNDSDDLVTNLRVQVNGSNVATPNCGAPSGFTVAAGNGSVTTTGWGTSWGVKFG 750
QY 703 GYNG-----NMSPNLKEGAKVNFKLKPNENN--- 728
Db 751 VGIVTDELOYDVGTLTEELIFGLGKTSRVDTRDLDFMRGCAFNSFAERAQVEMFKTT 810
QY 729 -----NTSKPLRFLAN-----ITATGGSVFDDIYANHSGRG---AELKMS 768
Db 811 TTAGDILSRQQTATLTANDATRVYGDVNPFTLTATMSGINADAYVNSQFNLDYQATASTT 870
QY 769 ENISNGAFTNSHVRGDDAFK-----INKDLTINATNSFSLRQTKDDFYDGYARNAI 823
Db 871 ATQASNVGQYAITGNANGSEYFSQRYQLVRQDGRLTVPQAOLIVSADAKTKVYGDADPTL 930
QY 824 NSTYIS-----ILGGNV--TLGQNSSSSITGNITIEK-----AANVTL----- 861
Db 931 --TYQVSGLKNSDTAAGVLSGNLGRVAGEN-----VGNIGILQGLGLNTANYTSLVGN 983
QY 862 EANNAPNOON--IRDRVIKL-----GSLVNGSLS--LTGENADIK 898
Db 984 DLRIITPAQLNVIADAKTKVYGDLPALTYQVSGLRKGTAGAVLNGGSLSRVAGENGVY 1043
QY 899 GNLTISESATFKGTRDTLTNTGTNFTNGTAEINTQCVKL-----GNVTNDGDLNITT 953
Db 1044 G---INQGLGLVSSNYTLNTOGN-----NLITIKALLNVIADAKTKVYGDADPALTY 1093
QY 954 HAKNRQSIIGGDIIN-----RKGSINIIFDSNDAEIOIGGNISQKEG 996
Db 1094 QVSGLKNGDGTAGAVLNGGSLSRVAGENGVYVINGQGLGLLSANYDLSYQ--GN----- 1145
QY 997 NLTISSDKINIKQITIKGIDGEDSSDATSANLITKTELK-----LT----- 1042
Db 1146 NLITIKALLNVIADAKTKVYGDADPS-----LTYQVSGLKNGDGTAGSILTGLNRA 1196
QY 1043 --EDLISGCFNAEITAKDGR--DLTIGNSDG-----NSGAEAKTYTFNNVKDSKI--- 1090
Db 1197 AGENGVYVINGQDLALNSGNYDLSYOGNLTITKALLNVIADAKTKVYGDADPSLTYQV 1256
QY 1091 ----SADGHNVTLN--SKVTSSNGGRESNDNDTGLTITAKNVEYNKDTISLKTVNIT 1144
Db 1257 SGLKNGDTAGAVLNGGLVVRVSGENVGNVYATQOGLGL-----VSGNYDL--AYQGNLT 1309
QY 1145 ASEKVTITTAGSTINATNKA--SIT-----TKGDTISGTI-----SGNTVSV----- 1184
Db 1310 ITKALLNVIADAKTKVYGDADPSLTYQVSGLKNGSDSGLTGLNRAAGENGVYGINQ 1369
QY 1185 -----SATVDLT-----TKSGSKI--EAKS---GEANVTSATGTIGTIGTISGNTVV 1225
Db 1370 GDALNSGNYDLSYOGNLTITKALLNVIADAKTKVYGDAD--PSLIYQVSGLKNGDTAGA 1428
QY 1226 TANAGDLTVNGAEIN--ATEGAATLTATNTLTTEAGSSITSTKQVDBLLA----- 1275
Db 1429 VLNGGGLVVRVSGENVGNVYATQOGLGLVSGNYDLAYOGNLTITKALLNVIADAKTKVY 1488
QY 1276 -----QNGSIAGSINAANVTLLTGTGLTITVAGSDTKATSGTLVINAKDAKLN 1322
Db 1489 DADPSLTYQVSGLKNGDTAGAV-----LN--GGSLSRVAGENV-----GVYINGQDLALN 1537
QY 1323 GDASGOSTEVNAVNASGOSVTAATSSSVNITGDLNT-----VNGLN----- 1364
Db 1538 ---SGN-----YDLSYQGNLTITKALLNVIADAKTKVYGDADPSLTYQVSGLKNGDTA 1588

QY 1365 --IISKDGRNTVRLRGKEIEVKYIQPG 1389
Db 1589 GAVL--NGGGLVVRVSGENVGNVYATQOQ 1613

RESULT 8

Q9HVN6
ID Q9HVN6 PRELIMINARY; PRT; 1417 AA.
AC Q9HVN6; 2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PA4541.
GN PA4541.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.F.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AE004867; AAG07929.1; -.
DR HSP; P22629; LSWC.
DR InterPro; IPR003880; Ppantne_attach.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1417 AA; 139958 MW; 212C916D5A55C39D CRC64;

Query Match

Best Local Similarity 21.9%; Score 636.5; DB 16; Length 1417;
Matches 352; Conservative 238; Mismatches 523; Indels 495; Gaps 72;

QY 1 MKKIYRLKSKRLNALVAVSELRGCDHTEKSGEPARKVRRHLAKPLSAMLISLGYT 60
Db 1 MNKSYTLVWQ-----ATGCWNVASEGTRRSK--SGRKAALVAVAGASLLGL-PC 47
QY 61 SIPOSVLASLGLOGMDVYHGTATMQ--VDGNKTLII-RNSVDAILNKKQFNIDQNMVQFLQ 117
Db 48 QAPAFALPS--GATVVSQDAGFQSTDRHVMVDDQSHKLITNNFESVRADERVSFHQ 104
QY 118 ENNSAVFNRTVSNQISQLKILDSNGQVFLNPNGITIGDKAIINTNGTFTASTLDSNE 177
Db 105 PGQDAVALNRVGRNGSDIOGRIDANGKVFVLPNGVVGKSAQVNVGGLVASTLDLADR 164
QY 178 NIKARNFTQTKDKALAEIVNHGLITVGKGSVNLGKVKNEGVISVNGSGISLLAQ 237
Db 165 DFLAGNYQFSGSG--ATVSNAGSLQASEGSIALLGARVSDGLIQAQLGDVALGAG 221
QY 238 KITSDIINPTITYSTIAAPENEAVALNGDIFAKGGNINRA-----ATIRNOGLSADSV 291
Db 222 GINLN-----FDGDLNLNQVDKGSVDLAHNGGLLRAD-- 255
QY 292 SKDKSGNIVLSAKGEAEIGGVISAQN-----QQAQGGKLMITGDKVTLTKTGAVIDLS 344
Db 256 ---GGVILMSARSADSLTKTVNNQGTLEARTLSRAGEGRIVLDGGEQGTVRVAGKQDAS 311
QY 345 GKEGETVLGDERGEGKNGIQLAK---KTSLEKGSITNVGKGGKGFALVWGDIALID 400
Db 312 AIGGN---GGLVLNQGAN--VEIORTAQVDTHADGAT-----CTWRILSHEVSVA 359
QY 401 -GNTNAQSGDIATGGFVETSGHDLFTKDAIYDAKEWLLDFDNVNSINAEPLFNNTGI 459

Db 1353 LILNGVSESIANVLGVAGSTVDIIGADATLTANNISGLF---GOYALAGNSKLTVASTN 1409
QY 402 NINAQ-----GSGDIATKGFVETSHD-----LFIKONAIVD-----AKEWLLDF 442
Db 1410 NLGASSVALAGAGDTLSLGSFNCTFGNSVTGSGVLQVTDAAEVTLTSSNGSVNAVIDI 1469
QY 443 DNVSINAED-PLFN-----NTGIND-----EFPFGTGEASDPKKNSELTLLTNTT 487
Db 1470 ADATLNLDDIALFNHLVTGNGLLVAKNDASTAFDFGTGGAFTGIVN-----LNTT 1523
QY 488 ISNLYKNAMTMINITASRKLTVNSSINIGSNHLLHSGKQGGGVQIDG----- 536
Db 1524 FALSADNAAL-ARATLKLSDSDSVTVGATDR-TLHGLDUNGTLIFDGPQPOANGV 1581
QY 537 ---DITSKGNNTIY-SGGWVDVH-----KNITL---DQGFNLITAAAFEGGNKARDA 585
Db 1582 TVTDLALNSGTISITGAGNENEPVTPPNVLEQDRGDILLELINAANYTGN-----A 1636
QY 586 ANAKIVAQGTVTITG-----EGKDFRANV-----SLNGTG----- 616
Db 1637 NLDDLVDGTATSTGTQGVESAIOQGGSTVANAHHNYGLTSSNGNGGSLYVNTLSALE 1696
QY 617 ---KGLN---IISVNNLTHNL---SGTINISGNITINQTRKNITSYNOTSHDSHWYSAL 668
Db 1697 LLANGANALLATESGLTANRVLNAELFGVGLVVDQAQNGALTLANGNNRYEGTIVTAG 1756
QY 669 NLETGANFTF-----IKYISSNSKGLTQYRSAGVNFNG-----VNGNM 708
Db 1757 ELILGANGAFGQTSLLNIASGASANINGYRTQVGAVTNSGAVTLNGGVLTSGLLTNGGI 1816
QY 709 -----SFNLKRGARVNF-----KLKPN-----ENMNTSKPLPIRELANITATG 746
Db 1817 LDLTGALNALAGSSVAVAGLTGAGTLNCGDLAVSATNSGLSQTHIADVASVTLTG 1876
QY 747 GGSVFEDIYANISGRGAELKMSKSEINISNGANFTLNSHVRGDDAFKINKDLTINATNSFS 806
Db 1877 TGTL-----GTSAVEVLGTNL-NGANAAMTNVLSGGVINTNAAVTLSCNNSFS 1926
QY 807 LRQ--TKDDEYDGVARN--AINSTYNTISILGNTVLTGGQNS-----SSITGN-ITIEKA 856
Db 1927 AHGIGTDELIVGAQSLGASSATVNLGTLTSHLILNGVSESIANVLGVAGSTVDIIG 1986
QY 857 ANYTEANAPNOONIRDRVIKIGSLVNGSLTGENADIKNLTISESATIFKGTROT 916
Db 1987 ADTALTANSG-----FLGQVALAGNSKLTVASTN---NLGASSVALAG-AGT 2032
QY 917 LNTGNTNCTABEINIT-QGVVKGNTVNDGDLNITP-----HAKNRQSI 962
Db 2033 LSLUG---FNGTFGNSVTGSGVLQ---VTDDAEVTLTSSNGVGTNVKVIADATLNDI 2086
QY 963 IGGD-IINKKGLSLNITDSNDAEIOIG-----GNISQKGNLTISSDKINTIKQITI 1013
Db 2087 ALFDHVLGTGNLNAVKNLATTAFFDGTGCGAFSGIVNLTTTFALSADNAALARATL 2146
QY 1014 KKGID-----GED-----SSSDATSNANLTIKTKEKLTE-DLSISGFN 1051
Db 2147 KLSDDSVTTVGTTRILHLGLDLNGTLIFDGPQPOANGVTVTDLALNSGTISITGAG 2206
QY 1052 KAP-----ITAKGRDLTI-----GNSDNGSGAEAKTVTFNVKDKSIS 1091
Db 2207 WNENEPVTPPNVLSLEQDRGDILLELIDADNVGTGANOLEMINGTTISAGOGVSTVQ 2266
QY 1092 AGHNVTLNSKVTSSNGGR-----ESND-----NDTGLTITAKRVEVN 1132
Db 2267 QGGYTVANATHNYGMTSNGSGSLYVNVYTLTSALELLADGANALLATESGLT---ANRELN 2323
QY 1133 KDTISLKTVNITASE-----KVTITAGSTTNATNKAISITTKGDISGTIS 1178
Db 2324 ALSGVLGVVDQAQNGALTLANGNNRYEGTIVTAGELILGANGAFGQTSLLNIASGA-S 2382
QY 1179 GNTVSATVDTLTKSGSKIEAKSGE-----ANVT-----SATGTIGGTI 1218
Db 2383 ANINGYRTQVGAVTNGTIVTLONGELTSTDTLINTGMINVTGDILNLENGGASSISGL 2442

QY 1219 SGTNVVNTANAGDLTV-----GNCAINATEGAATLTATGNTL---TTEAGSSITSTKQV 1271
Db 2443 TNGI-LNTKGDFTFISDNNGLAQTNISDGSVTLGNGGTIGTGNLGSVDVLGDL 2501
QY 1272 DLLAONGSIA-----GSIN-AANVTLN-----TTGTLTTVAGSDIKATSGT- 1311
Db 2502 NLVADN-SLANVISGDCGTINTTATVTLSONSFSGAHQICTNGELTVGOANUGASSATV 2560
QY 1312 -----LVINAKDAK-----LNGDA-----SGDSTEYNAVNA-----SGSGSVT 1344
Db 2561 NLGTLTSHLILNGVSESIANVLGVAGSTVDIIGGADALTANNISGLQVALAGNSKLT 2620
QY 1345 AAT-----SSSVNI---TGDLNTVNCIN-----IIISKDG-RNT 1373
Db 2621 VASTNNLGAASSVALAGTGTLSLGSFNCTFGNSVTGSGVLQVTDAAEVTLTSSNGV 2680
QY 1374 VLRLGKEIYKIQPGVASVEEVIEAKRVLEKVKDLSDERETLAKLGVSAVREVPNNT 1433
Db 2681 VTIDIADATLNL--DDIALFNHALTNGLLNVAKNKNDASTAFDFGATVGGATGTVNLNS 2738
QY 1434 ITVNTQNEETTPSSQVILISEGKACFSSGNGCARVCTNVADDG 1475
Db 2739 TFDLSGNTTVLAQATLKLSSGNLT-SVNGGVQNIPTLAMNG 2779

RESULT 10

Q9F285 ID Q9F285 PRELIMINARY; PRT: 3705 AA.
AC Q9F285;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Yaph protein.
GN YAPH.
OS Versinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 BIOVAR ORIENTALIS;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277631; CAC14227.1; -.
DR HSSP; P22629; ISWG.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01656; VACCYTOTOXIN.
SQ SEQUENCE 3705 AA; 370756 MW; 714FDF16455968C9 CRC64;

Query Match 8.0%; Score 594; DB 2; Length 3705;

Best Local Similarity 21.5%; Pred. No. 1 4e-11;

Matches 409; Conservative 262; Mismatches 695; Indels 538; Gaps 91;

QY 14 NALVAVSELA--RGCDHSTKESG---EKPMKVRHLKPLKPLSLGVTSIPQSVLA 68
Db 972 SGVTVTDLALNSGTVNIITGSGWSDMTDPLATNVSTLEQDRAGSTLELINATVTDGIDA 1031
QY 69 SGLQGMVDVHGTATWQVDGNKTIIRNSVDAILNMKQFNIDQNEVQFLOENNSAVENR 127
Db 1032 LDL----LVNGTAITSGTGVSQSAIQGGSTVA-----NAHHNYGLASSNSGDSGLVYN 1082
QY 128 VTSNQISQLGILDSNGQVFLINPNGITICKDAIN----- 164
Db 1083 VT---LSALELLADGADALLATESGLTANR--VLNAELFGVGLVVDQAQNGALTLANGS 1137
QY 165 NCFSTASTLDTSNENKARNFTFEQTK-----DKALAEIVNHGLITVKGDK 209
Db 1138 NRYEGTIVTAGELILGANGAFGQTSLLDIASGASANINGYSQTVGAVTVNGVTVLGSGG 1197
QY 210 SVNLIIGKVKNEGVISVNGSGSISLAGQKITISDIINPTITYSTAAPENAVNLGDIKAF 269

Db 3 NKRYKLPKVKKCLVPVAENIKSASGSSSSSKIAEDQEEBPDLSIACSLPLSSSIH 62
 QY 45 LAL-----KPLSAMLISLGVTSIP-QSVLASGLQG-MDVHGTATMQVD----- 86
 Db 63 LGLHNHSPKLVKFKGKNSLVLLSL-----MPAKVWADSSNAIVDHSHGAKQFAVDERDPK 118
 QY 87 -GNKTIIRNSVDAIN-----WKQFNI-----DQEMVQVLOENN 120
 Db 119 NKEKQV-----VINIAKPDQOISDNHSEKFNIPNSAVFNNSIKEGNSQLVGLGENK 172
 QY 121 N-----SAVENRVTSNOLSKILDSNGO--VELINPAGITIGKDALINTNGFTAST 171
 Db 173 NLGSOAKTIFNQVTCDOESKISGLEVEGEKADLIIFINPGLVINGVKTINTDFEAST 232
 QY 172 LDIENIKARNETFEQTKKALAEIVNHGLITVGKDG-----SVNLIGKVKNECVI 224
 Db 233 SEVVEPHIKQLN-----VQRKVLIIGKGVATNGLSHFVVAKNIBQOKV 278
 QY 225 SVNGGS-----ISLLAGKITISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRA 277
 Db 279 SIEGDKPAKLANVFAAG-----NLTYD-----VNRDQV-----NRNT-- 312
 QY 278 ATIRNOCKLSADSVKSGNIVLSAKEGEABEIGGVISAQNOQAKGKLMITGDKVTLKT 337
 Db 313 ----NPKKPTDTRKD--NIAIS--GE-----SAGSMYGRNRIKFIIV----- 347
 QY 338 GAVIDLSGEGGETVLGGDERGCKN--GIOLAK-----KTSLEK--GSTINVS 383
 Db 348 -----DKGAGVNHQGVIFAEDDINILTDGNSRLNKVYADVVRWG 388
 QY 384 KEGGFAIVWGDIALI-DGNINAQGGDIATKGGFVETSGHDLFIKDAIVDAKEWLLDF 442
 Db 389 K-----DIELANNGQIHADQOILNATGHVKLNDGSSVISNNLIGISALN--LTL 436
 QY 443 DNYSINAEDPLF--NNTGINDPEFTGTGEASDPKKNSEL-----KTLT--TWTTI 488
 Db 437 ENATVSANLNSFRVNDTKLNNLSKVSARAADLOGNLDKASVLAHKLTLINISNDVSL 496
 QY 489 SNVLYK-NAWMTNITASKRLTVNS-----INIGSNHLLHLSKQORG-------V 532
 Db 497 NNGSKLSANLKIKKVRLDLNLSSELSANLTLNNTITLKNKSKFTAGNLTNNTVNV 556
 QY 533 QIDGDTISGKNLITISGGMVHVKNITLDOGF-----LNIT-----AAS 572
 Db 557 TLNDSLEAANNLTL-----NVTKNVLNDSKLSANKLIDLNVDTNLTLSKSTLSAGE 610
 QY 573 VAFEG-----GNKARDAANAKIVAQGTVTITGEK-----DFRANNVSLG--TG 616
 Db 611 LTFKVKVNTLNDSLEAANNLSNASHNVTNNKSKLSAQKADIKAVNLTLNDDTTLTA 670
 QY 617 KGLNISSVNNLTHNLSTI-NISGNITINOTTRK-----NTSYWOTSHDSHW-NV 665
 Db 671 KNLDI-----NSTITNGTIAGIFANITTEKLNKKEKALILAEQLNF--TVNGSHYENK 724
 QY 666 SALNLTGANFTFK-----YISSNSKGLTQYRSAGVNFNGVN-----GNMSPNKE 714
 Db 725 GDVSKOKATVTSKNSDFTSGSKLVNAQNLKVVYNNFTISQGGDITLIGNVTLN-AS 783
 QY 715 GAKVNFKLKPNENNTSKPLRFLANITATGGGVFEEDYANHSR-----GAEKWSI 770
 Db 784 GTFIN-----SGNLTVTKLIDVGDQNTFNKGNLTVEGDLHKSRTKINDGKLISIKL 838
 QY 771 NISNGANTLNSHVGDGDAFKI-----NKDLTINATNS--NFSLRQTKDDFYDGYARN 821
 Db 839 NISSEADFINNGTLIGTEALKIATKGNFTNKEKAILASLLDISVAEGKKTENN-- 894
 QY 822 ANSYNINISILGGNVTLLGGSSSSITGNITIEKAANVTLEANNAPQOINRVRVILKGS 881
 Db 895 TIESCKNLNINTGAFNLVDNATIRSGVLNITSGNV-----SNN-----GT 937
 QY 882 LTVNGSLGTGENADIKGNLTISESATFKGTRDRLNITGNFTN--NGTAEN-IPQGVV 938
 Db 938 LISNERLNTS-----AANFTNESNGTWSNGLNIIA 970

QY 939 KLGNTVNDG-----DLNITTHAKR--NORSIIGG-DIINKKGLSLNITDSN---NDAIQ 986
 Db 971 KQGNITNKLIAIROQLNLTAVADNITNDSNISKIAVLHSLGNSLKQDQVYNLGEIY 1030
 QY 987 IGGNISOKENLTISSDKINITKOITIKK-----GIDGEDSSSDATSNAN 1031
 Db 1031 AGNISVKAHQL---KNQVKLMGDITTKTKEGOASYKLYQASNGGHFGDGS--GYSBGD 1086
 QY 1032 LTIKTKELITLEDLSIGSFNKAETAKDRDLTIGNSNDGNSGAETVTFNNVRSKIS 1091
 Db 1087 LNIKGFADLNDKLTQVQIRG--IYA--GRLTFNKSNAAGK-----SEII 1128
 QY 1092 ADGHNVTLASKVKTSSSNGGREGSNDNDTGLTIFAKNVENVKDIOTSLTKTVNITASEKVT 1151
 Db 1129 NRG---TINVKNKLSYDSVFNENMQSOKVDLYTKIPEAKSDIE--LTFKTNHTHPVL 1183
 QY 1152 TAGSTINATNGKASITTKTGDISGTISGTVSVSATVDLTT-KSGSKIEAKSGEANVTS 1210
 Db 1184 NFKSNNEKKYRNSNTKNFKSIGDLINEALSDSAPAEIAYYSGS-----SSNINPVSY 1239
 QY 1211 TGTIGTISGTVNVTANAGDLTVNG--AEINATEGAANTLTATGNTLTTPAGSSITSTK 1268
 Db 1240 LAALGNANSSPHYLNTALKHLIGNGWQDDLLKQENIKVLKQWEDFKKDKGAS----- 1294
 QY 1269 GOVDLLAONGSTAGSINAANVTLTGTLTVAGSDIKATSGTLVINAKDAKLN-GDASG 1327
 Db 1295 KMLDLYPNTDKAKI-FAGIIRNGNDTISDVESEDFKKYSKFQ-NGEAKNODTGTDSY 1352
 QY 1328 DSTEVA-----AVNAGSGSVTAATSSSVNITGDLTVNGNLNISKDGRNVRIRGK--E 1380
 Db 1353 DSTKASEKVKVENVDHKENI---DBHKLNIKGHEITVPGVSPENLNKNDHDPDKLGE 1409
 QY 1381 IEVYIQPOVASVEEVTEAKRVLEKYKD--LSDEERETLAKLGVSAVRFPENN 1432
 Db 1410 IDKSIISELLA---QPVYTEKSAARSDPRVQNDKEALDNLVRLTSLYINQNN 1460

RESULT 12
 Q8XPU7

ID Q8XPU7 PRELIMINARY; PRT; 2737 AA.
 AC Q8XPU7;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Probable hemagglutinin-related protein.
 GN RSP1539 OR RS05770
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG plasmid megaplasmid.
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salenouat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646085; CAD18690.1; -;
 DR InterPro; IPR000566; Lipoclin_CyLFABP.
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 2737 AA; 268490 MW; 12F383C02C101D0C CRC64;

Query Match 7.6%; Score 561; DB 16; Length 2737;
 Best Local Similarity 22.4%; Pred. No. 1.2e-10;
 Matches 326; Conservative 187; Mismatches 546; Indels 396; Gaps 64;

QY 82 TMOVDGNKTIIRNSVDALINNKOFNIDONEMVQLOENNNSAVENRVTSTNOISOLKG--- 138
DB 155 SLTVDGVGLIINNSLAGGGTFLGNGVGN--ANLAASGPASTILITQVTGTEPIRINGTVE 212
QY 139 ILDSNGQVFLNPNRGTITIGDAINTNGFTASTLIDISNENIKARNFTEQTKKALAEIV 198
DB 213 VFGSPASVIFAAPAGVYTOGAGFTNPTRVTLSTGTPOFLANGSGANVAFDQA--TAVGFLV 270
QY 199 NHGLTIV-----CKDGSVNLIGGKVRKNEGVISVNGGSIISLLAGQKTIIDII 245
DB 271 NSGRIOIEPAAGSTAGAGIEGTGVAINGLIGTV-----GVNAPLEAGNQINV--IA 319
QY 246 NPTIYSTAAPENEAENVLGDIFAKGGNINVRAATIRN-----QGLK--SA 288
DB 320 NQOQVAPVATGTGRAGSDQWVSGAGANAAANSQAQNGLAIDATAFGAMTAGQIKLISTA 379
QY 289 DSVSKDKSNIVLSAKEGEAEIGGVISAQN--OQAGGKLMITGDKVTLKTGAV--ID 342
DB 380 QGLGYRAAGDLAANTSNVNDANGDVGVNGVYGOOTAG--ITTTGSVST--SGAVRAQOD 435
QY 343 LSGKEGGETYLGDERGEGKNGIQLAKTSLKSTINVSKEKG--GFAIWMGDIALID 400
DB 436 VTIGAGGDDVTLGG-----AAQAGNVTVSAGNVAGSGDLAAVHGLSVSAGNSANLG 487
QY 401 GNINA-----QGSQDI-----AKTGGFVETSCHDLFIKDNAIVDAKEMLLDFDN 444
DB 488 GNLNAASIAVTAQGGKDGTDITLGGKVASPGGIALNAARDTTIAGO-----LTGSG 539
QY 445 VSINAEDPLFNNTGINDFPTGTGASDPKKNSELKTLTNTTISNYLKNAWTMNITASR 504
DB 540 VSVNAGRNLVSAGLIG-----SVGDLNLAAAGSVSTTGAVTT-----QANLTASA 585
QY 505 KITVNSSTNIGNSHLILHSKQORGQVQIDGDIITSGKGNLTIIYSGWVDVYHKNITLDQ- 563
DB 586 QDVRLLGGTTAANGHVAIQANA-----GSITT-AGLTAAQDIALTAGQATLGAA 635
QY 564 ----GFLNITAASVAFEGGNNKARDAANAIVAOQTGTVITTEGKDFRANNVSLNGTKGL 619
DB 636 TOTGNGTLVTAGNTVGGAGNTASKAID--VQAGGSVDVSG--NVSANRIAMQAAGR-- 688
QY 620 NIISVNNLTNLSGTINISGNITITNOTTRKNTSYWQISHDSHNVVSALNLETGANFTFI 679
DB 689 ----DGVGDI--RLGGNVGAPGTITLN--AARDTTIAGSVVSDSLN----- 727
QY 680 KYISSNSKGLTTOYRSSAGVNFNGVNGNMSFNLKEGAKVNFKLKPNENNMTSKPLPIRFL 739
DB 728 -----LATQRLNSVGVVGVSTKGNVSLTARTGA----- 755
QY 740 ANITATGGGVFFDIYANHSRGAEELKMEINISGANFTLNHVRGDDAFKINKDLIN 799
DB 756 --VTT-----QCAVITPGNLVSVSGADTSLGGQVSAAGTASVN----- 791
QY 800 ATNSNFSLRQTKDDFYDYARNAINSTYNISILGONVTLGGQSSSSITGNITIEKAANV 859
DB 792 -----AGNLSLTAGOISGN--GTLTLNAGONL 816
QY 860 TLEANNAPNQOINRDIRVILKGS--LLVNGSLSTGENADIKGNLTISESATEFKGKTRDTLN 918
DB 817 TIGGQVQVSG----ADANLQAGSNVTYNGALTSTG--NASVLAGQSGIA----- 857
QY 919 ITGNETNNGTAEINITQGVVVLGNVTNDGDLNITTHAKRNQRSI--IGGDIINKKGSNLT 977
DB 858 LAGDVAAGGNATLDAQ-----TITGPNLSAAQATAKVTGGSIDLGGQVKQKQVALTAN 911
QY 978 DSNDIAEIQGNISQEGNLTISDPKINITKOITI--KGIIDGEO-----SS 1023
DB 912 GSNLGDVRLGGVAG--APGSVTLISA-----TRDATLGSNAIAGGDLTATAGRLNVNGNA 965
QY 1024 SDATSNANLTIKYKELKLFEDLSISGFNKAEITAKDGRDLITGNS--NDNGSGAAK--- 1078
DB 966 ASVNGNVNLTAAQAGLQASTGSIQA---NOGDVNAVTAQOGLNVGSGVYAGRNAGLVAQGN 1022

QY 1079 -TVTFNNVSKIS-ADGHNVTLNSKVKT-----SSSN-----GGR----- 1112
DB 1023 ATVSGNLTSLKASISGGNTLTLGOLTKVGGDLQASAAANALSVQALNVYGGNAILRGTDI 1082
QY 1113 -----ENSDNDT-----GLTTAKNVEVKN-----DITSLKTVNITASEKVTTPAG 1154
DB 1083 AVGSAAGQSNVAGOTGLDVAASRGLTL-AGNSNANALNGGATIVNOGSTLATORATVSG 1141
QY 1155 STINATNGKAS-ITTKTGDID--SGTISGNTVSATVDLTTKSGSIEAKSGEANVTSAAT 1211
DB 1142 MVTNAGMLAANQLTVSATDLVNRGTVGGQAVKLNTTGNLDNAGGLVVGSTOLDVTAGALT 1201
QY 1212 GTIGGTISGN-----TVNVTANA-----GDLTVNGGAEIN-----ATE 1244
DB 1202 SNRGGTFFGDLGKSPPTTGNLTFTVNGGAGSPNAGGOLLAGNLTNTPTNOAFDPSAA 1261
QY 1245 GAATLTATGNTLTTEAGSSITSTKGOVDLLAQNGSIAAGSINAA-----NVTL 1291
DB 1262 TACTLNA-NNTLTL-----SIOSINNTGTWNVOGSSVA--INAAQGFISGCTIOKAGNLSL 1314
QY 1292 NTTGTLTT-----VAGSDIKATGTLVINAKDAKINGDASGDSTEVNAVNASGSGVTAAT 1347
DB 1315 STAGALANSOIVGGSNNVALSAGTLT-NTGTIHADGNLAGNVRNAGTAEALGNI-AVT 1372
QY 1348 SSSVNTIGDLNTVNG 1362
DB 1373 GSNYDNOGGKTOANG 1387

RESULT 13
Q9HWU6 PRELIMINARY; PRT; 1018 AA.
AC Q9HWU6
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Probable adhesin.
GN PA4082.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen."
RL Nature 406:959-964 (2000).
DR EMBL; AE004824; AAG07469.1; -.
KW Complete proteome.
SQ SEQUENCE 1018 AA; 100433 MW; F373734D77FFA94D CRC64;

Query Match 7.5%; Score 553.5; DB 16; Length 1018;
Best Local Similarity 22.6%; Pred. No. 7.4e-11;
Matches 301; Conservative 181; Mismatches 453; Indels 399; Gaps 61;

QY 1 MNKYLKFSKRLNALVAVSELARGCDHSTKEGSE--KPARMKVRHLALPLSLALLSL 57
DB 1 MNKVAL-----VNVVSGCMNVVSESGRRRRKGPAGAKA-----AIAVLALL 43
QY 58 GVTST-POSVLASGLQGMVVGHTATMOV---DGNKTIIRNSVDAIL--NWKOFNIDONEM 112
DB 44 GATALAPAYALPS---GGTVYGGSGANGEIHLSCGNSLVNQVKLLIANWDSFVAAGER 100
QY 113 VOFLQENNSAVENRVTSTNOISOLKGLDNGOVFLINPGNITIGKDAINTNGFTASTL 172

Db 1123 LTVTGSNTVATG-----LGPDEK-----GTOIA-----GMTTGAAG-- 115A

Query Match 7.4%; Score 548; DB 16; Length 3930;
Best Local Similarity 21.9%; Pred. No. 4.8e-10;
Matches 366; Conservative 227; Mismatches 641; Indels 441; Gaps 82;

QY 386 KGGFAIVGDIALLDIGNINAQSG-DIAKT-GGFVETSGHDLF-IKDNAIVDAKEWLLDF 442
Db 1155 -----VNESSVA-VDSGGGVGTGIAISNTSGGNVTEGGVDLFRISGTAL-----SL 1200
QY 443 DNVSINAEEDPLFNWGTINDEPFTGTGEASDPKKNSELKTLTITNTTISNYLKNWMTNITA 502
Db 1201 DNASGCIYS--FNQTKVDIV-VGGPG-----FLVNSAATVGNLVTNNI-- 1245
QY 503 SRKLTVNSINIGNSHLILHSKQORGGOVIDGITSKG--GNLTIVYSGGWVDVHKNTIL 561
Db 1246 -----AGADVSLTNT-----GTIGIGTITNSGTGNGVGVVSGSAAI--TVSA 1287
QY 562 DOGLFNITAAVAFEGGNKNKARDANAAKIVAOQTVTITGEGKDFRANVSLNGTGLGNI 621
Db 1288 D-----ISSSATAPGTAVKVDGTGTGGSVTFPSGLITSTGTGTGVSVSNTAA--GSGVGFGA 1340
QY 622 ISSVNLNLHSLGNTISGNITINQTR-----KNTSYWQTSKSHDHWNSALNLET 672
Db 1341 V-TVSGAAGNGIGISGNAGSVTFGTNTVTLGSAANAAGINFSGTNADVLFGTINVTMGA 1399
QY 673 GANFFIKYISSNSK--GLT-----TOYRSAGVNEGVNGVNMSENLKEGAKVNFKLKP 725
Db 1400 GANQTGIDFGSSATAGFLTTITGTDLTSRGIDLSSTTGKNTITFARGSSIT----- 1453
QY 726 ENMTSKPLPIRFLANITATCGSVFFDIYANHS--GRCAELKMEISINSGANFTLSHV 784
Db 1454 -NVG-----VGVLSGGTITAYSANANFTFGD-----NAGDGLQSTISAAA 1494
QY 785 RGDPAFKINKDLTATNATNSFSLKOTKDDFYDGYAR--NAINSTYNSIILGNVTLGQON 842
Db 1495 GGYVTNTIGLDTPLG--NYNFN-----DVFTGSAHLASAVGGTIVMVSQMGVIAAGTDG 1547
QY 843 SSSSITGNITIEKAANV-----TLEANNAPNOQINRDRVVKLG-----SLVINGSL- 888
Db 1548 LSAAVT-TISVAQADALAGTGTAFVGTGTVLDSSTPFTLDSQOSTGTGNNNTSIVSGTIQ 1606
QY 889 -----SLTGENADIKG-----NLTISEATEPKGTROTTLN-----TGN 922
Db 1607 PNVVGNLGSAGGNVTNEGVTGTGDFLHLLGSNOVRNTAFNFSGASGVNFVDQSAGG 1666
QY 923 FTNNGTAEI-----NITQGVVVLGNVTNDGDLNITTHAKRNOBSIIGDDIINKKSL-N 975
Db 1667 FNNAGGIVIEGTVSNVATGTAFKVLGDLNLSIT-----NNNVNAGTLLDVNGGTGN 1721
QY 976 IT-----DSNNDAAEIOIGG-NISOREGNTLISDKINIKQITIKKIGDGEDSSDAT 1027
Db 1722 ILLRGTLPNATPGLTGGISIANRSGGL-----VNFTDKVIV-----GGAGVSLT 1769
QY 1028 SNANLTIKTKELKLTEDLSIGFNKAETAKDGRDLTIGNSDNGNSGAEK-----TVTFNN 1084
Db 1770 GNTGTVTTFADL-----DITTSGATAFASAGGGTVNVTGTGINATNAQAAALDGTAGINF 1825
QY 1085 VKDSKISADGHNVTLNSKVKTSNGGSGRESNDNT-----GLUTIT 1125
Db 1826 ASTSATFASNGIDQLNLSGTFSGTGTTLNTGSGTGFNFVSGSATNLSGGNAVISYGGTIA 1885
QY 1126 AKNVEVNRADITSLKTNITASEKVT-----TTAGSTINAT--NGKASITTKTG----- 1171
Db 1886 SNGTGAASVIOELTGGSTVLSGNVTDGLAGAGGNVTVTGIDNGTAATVTFSGSKQIDSG 1945
QY 1172 -----DISGTTISGNTVSVSAVDLTTSKSKIEAKSGEANTVTSATGTI----- 1216
Db 1946 ATDGVSLLGNPNGLTAFISGLVITTSAGAGFHAS-----TFGTGIVTVTKDGGANNTI 1999
QY 1217 -TISANTV--NVNANAGDLT-----VNGAEINATEGAAT-----LTATGNT 1255
Db 2000 TTTTGNALNDNVTVGVGINFDSISSNGTGTGIALNNVSGGAINLGAVDLEGITSRQVD 2059
QY 1256 LTTEAGSSITSKQGVDDLLAONGSTAGSINAA-----NVTLN-----TTGLTITVAGSDIK 1306
Db 2060 VSGTLGSLTFTSLNIGLNAAN-AIGDLNSAALGVSNITAGDFDVGGSFAGTIGDMA 2118
QY 1307 ATSGTLVINAKDAKLNGDASGDSTEVN-----AVNAS--GSGSVTAATS----- 1348

Db 2119 GTTGTGTIQLGDTVNNNPAGQSTIANVGYGVQFSSATNAQLVFGDGAGPAESSIATTTGG 2178
QY 1349 -----SSNYITGDLNTVNLNII-----SKDG--RNTVLRGKEIB- 1382
Db 2179 OVTHATDTTPTSGDYFNDFNSGDTSNLSAVRVYVTVAGTGDGSLANPGSYLGAQAST 2238
QY 1383 ---VKYIQGVASVEVIEAKRVLEKVD-----LSDEERTFLAKGV-----SAV 1425
Db 2239 ANVVVLIDKNVGAQITDLSGTTFNLDGQVLLAFKSGDAADVDSQVLTGSGGASAAF 2298
QY 1426 REVEPNNTITVNTQNEF--TTRP-----SSOVI--ISEGKACFSSGNGGARVCTNV 1471
Db 2299 HFTTQNTPTIISAPGGIDTLRPVLQSNNTATSVINFATSGTGTGTGIIENLIVSNV 2353
RESULT 15
Q8RDO9 PRELIMINARY; PRT: 3165 AA.
AC Q8RDO9;
DT 01-JUN-2002 (TREMREL. 21, Created)
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Fusobacterium outer membrane protein family.
GN FN1449.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 25586;
RC MEDLINE=21886394; PubMed=11889109;
RX Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen M., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010650: AAL95642.1; -;
KW Complete proteome.
SQ SEQUENCE 3165 AA; 331849 MW; 6DFE562C8490DEAL CRC64;

Query Match 7.3%; Score 543; DB 16; Length 3165;
Best Local Similarity 22.5%; Pred. No. 5.6e-10;
Matches 372; Conservative 222; Mismatches 535; Indels 524; Gaps 85;
QY 68 ASGLQGMVVHGTATMOVDGKNTII-----RNSVDAIINMKQFNIDQNMVQFLOENNSA 123
Db 396 AVGLYSADPTKFNNTFKIESGKTLDELKGNSTFGLLGNNT--TVTNSPLLSKYLNNHTSD 454
QY 124 VFNRTVSNQ-----ISQLGILD-----SNGQVF-----LINPNG-----ITIGKDAIIN 163
Db 455 KINIVSFGEGASLFYATSKAKAILDEDYKVTNGDAISTAVLVANNGANVEIASGKKLETN 514
QY 164 TN-GFTASTIDISNENIKARNFTFEQTK-DKAL-----AEIVNHGLITVKGDSVNLIG 215
Db 515 TNAGLIAINGTVGFTSVAKNNGTLLSTRIDKGIIVTSAANGENSIGTITMKNKNAVGLIG 574
QY 216 GK--VKNEGVISVNGSGISLLAGKITTIDINPITTSIAAPEANVNLGIFAQGN 272
Db 575 SKGSNLKNTGKIELEAVS---SAGVVAEDSNMTNSGTTSEIIV--NKEASVG-IYAKETS 628
QY 273 INVRAATIRNOGLKLSADSVSKSGNIVLSAKEGEAFI-----GGVISAQNOQAKGKLM 327
Db 629 ISSVKNVKNKKEGKIEKADGDKRSAGI-YSKKEGGAARLTIENTGNTIEVAKKASA---GIY 684
QY 328 ITGDKVTLTKGAVDILSG--KEGETVYG--GDE-----RGEKNGIQIAKTSLEKSGT 378
Db 685 TKNESTQANTQSEVTNSGLVKMSAENSIGTGMGEKSKITNTGTGKTKGIEIVEK----- 736
QY 379 INVSGKEKGGFAIVWGDIALIDGNINAQSGSDIAKTGTFVETSGHDLFIKDNAIVDAKEW 438

Db 737 -----KSAGILVTNESAVIN-----SGRISLNSISSSSDGL----- 769
Qy 439 LLDENVSINAEPLFNNTGTINDEFPPTGTGEASDPKKNSELKTTTLTNTTISNLYKNAMTW 498
Db 770 -----VGISVDG-----SSTGEND----- 783
Qy 499 NITASRKLTIVNSSINIG-----SNSHLLHSGQORGGVQIDGDTSSKGGNLTIVSG 550
Db 784 ---ASGEIKVDAAYSTGMLSSGGDVTNAGKIALEKKESVGMVATNANVTNSG---AIPKG 837
Qy 551 CWDVHKKNITLDQGLNITAASVAFEGGNKARDAANAKIIVAQGTVTITGEGKDFRANV 610
Db 838 IFIDKESVGI---YKINSSSI-----ANKVTNSGTIDDIAGTSKTSAGIY 882
Qy 611 SL--NGTGKGLNISSVNNLTHNLSGTINISGNTIINOTTRKNTSYWQTSKSHSNVSA- 667
Db 883 SIIENGATKTL---SAVNN-----GNITING--KKSIVIYAKNESSHANTESD 925
Qy 668 -----LNLETGANFTFIKY---ISSNSKGLTQYRSSAGVNFNGVNGNMFSNFKEG 715
Db 926 VTNSGKIEVKNESGAVLAEKYKVTNCGSGTNGIIVSAKSA-----GIIGKLGSEI--- 977
Qy 716 AKVNFKLKPNB-----NMNTSKPLPIRFLANITATGGGVFFDIYANHS-- 759
Db 978 --INSIIKTETATPTVATDGVGVLNSNK-----ATNTSGGVITLD--TNYSTG 1024
Qy 760 --GRGAELKSEINISGANFTLNSHVRGDDAFKINKD-LTIN-----ATNSNFSLR 808
Db 1025 MYGEANSQLTNEGNT-GTNKEYIVGMAGDSSTVTNKNITLNGKKATGIFGKNSSLLN 1083
Qy 809 QT-----KDDFYDYARNAINSTYNISILGGNVTLGG----- 840
Db 1084 ETTGKIPTKEESVGMYSLSLKNATNGITTEKTSAGMLGDKANIENDSSIPTKEEMS 1143
Qy 841 -----QNSSSSIT--GNITTEK--AANVTLEANNAPQOINRDRVIKLGSLLVNGSLSLT 891
Db 1144 AGMVKNGTSKATNKGTVTTEKTSAGILAEIDEA-----NGG-TVS 1184
Qy 892 GENADIKGNLTISP--SATPKGTRDTLNTGNFTNNGTAEINITQGVVKGNTVNDGDL 949
Db 1185 GLN-ETTGTITVSETSAGMLGVKSA-----VTASTAKL-SLTNKKDI 1226
Qy 950 NITTHAKR-----NORSIIIGD--IINKGSLNITDSNDAEIOIG-----G 989
Db 1227 NINTKNSAGIMVYNVESTAVKENVLAENTGTINTLTSSATNEKNIGILANKATGINTGNI 1286
Qy 990 NISOKE--GNLTISSDKINTKQITI--KKGID--GEDSSSDATSNANLTIKTEL--KL 1041
Db 1287 NINSKESIGMLGQNASITNNKTITLSGKIGMLSKDTSIADNDIINNVCKESLGML 1346
Qy 1042 TEDLSISGFNKA-EITAKDG-----RDLTIGN-----SND 1070
Db 1347 GEDSGTVKNKNTISVTAEGVGFVRDNGVKGSGTGENTSTGTITILENKEAVGIFAKNN 1406
Qy 1071 GNSGAEAKTVTEN-----NVKDSKISADGHNVTLNSK-- 1102
Db 1407 GTSDSAKNSTIINLGKADGSTIKESLIGMPAQAEAGKANKVNTK-----DINVTNKS 1460
Qy 1103 -----VKTSSN-----GGRESNDNDTGL-----TITAKNVEVNDKITS 1138
Db 1461 VGIYAKNDASNITDVLNENTGDNINSKESAGVYAPKANISKVGTITLKN---SIDSNGS 1517
Qy 1139 KTVNITASEKVTTAGSTINA-----TNGRASITTKTGDISGTISGNTVSV-- 1184
Db 1518 SAVYVSGKGVKANTAGVKINLGTVQNRVAYVYVNGKDS--ALAGADIGKITGVGVYLO 1575
Qy 1185 -----SATVDLTTKSGSKIE-----AKSEANVTSATGTIGTISGNTVNV 1226
Db 1576 GTSGDKRATLD---KNTSKLDYTLQGTGNGLIIGLLKGETNIQSYT---KGIVGNTVAAT 1629
Qy 1227 --ANAGDLTVNGGAEINATEGAATLTATGNTLTTEAGSSITSTKGOVDLIAQ----- 1276

Db 1630 SPSPKAKVAIGIYADAQGTVG-----TPYNTT-----PITAGKSGVGIFADKDSNINYT 1679
Qy 1277 -----NGSIAGSINAANVTNTTGLTTLTVAGSDI--KATSGTLVINAKDAKINGDASGD 1328
Db 1680 GNMEIGDGTACTGTFITKKIGATGCKVTLGTNTIKLGTGKVVAIAISEGTTFNG---GN 1736
Qy 1329 ST-EVNAVNASGSGSVTAATSSSVNI---TGDLTNTVNGLINIISKDGRNTRVLRGREIEVK 1384
Db 1737 ATIELVGSNIQGVG-VYAKKSGSTVNIDHWTFFNNNGNSAEVESEBGRVYIN-ANKNLKPK 1794
Qy 1385 YIOPGVAASVEEVIEAKRVLEKVKDLSDEERETL 1417
Db 1795 MVLTHVINGETSIATGKTVTVSNDGSIITAKENI 1827

Search completed: March 24, 2003, 15:24:13

Job time : 86 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:23:05 ; Search time 26 Seconds
(without alignments)
5461.175 Million cell updates/sec

Title: US-10-092-880-4

Perfect score: 7407

Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	7327	98.9	1477	2 B43855	high-molecular-wei
2	4957	66.9	1536	2 A43855	high-molecular-wei
3	876	11.8	1910	2 AF0394	probable adhesin h
4	744.5	10.1	2154	2 F83068	hypothetical prote
5	636.5	8.6	1477	2 A83080	hypothetical prote
6	602	8.1	3705	2 AD0123	probable autotrans
7	593.5	8.0	4919	2 T31105	hypothetical prote
8	593.5	7.5	1577	2 A35140	hemolysin A precu
9	553.5	7.5	1018	2 H83135	probable adhesin P
10	531.5	7.2	1268	2 B99789	hemagglutinin/hemo
11	531.5	7.2	1270	2 E85649	hypothetical prote
12	523.5	7.1	4152	2 T31102	filamentous hemagg
13	520	7.0	1975	2 B81192	hemagglutinin/hemo
14	520	7.0	2535	2 AC0304	probable hemolysin
15	517.5	7.0	1995	2 G81044	hemagglutinin/hemo
16	517.5	7.0	2020	2 C48399	ABC-type transport
17	512.5	6.9	2340	2 B71704	cell surface antig
18	511	6.9	2249	2 A41477	190K surface anti
19	508	6.9	2015	2 B81989	hypothetical prote
20	504	6.8	3295	2 AE0074	probable adhesin Y
21	486.5	6.6	928	2 S54699	hemolysin-heme com
22	485	6.5	2059	2 D82671	surface protein XF
23	480	6.5	1594	2 T43072	hemolysin A Edwa
24	477.5	6.4	1749	2 AG2560	hypothetical prote
25	475.5	6.4	1487	2 S75138	hypothetical prote
26	475.5	6.4	2021	2 A97859	190-kDa cell surfa
27	474	6.4	5291	2 F90696	hypothetical prote
28	472.5	6.4	1152	2 AE1852	hypothetical prote
29	471	6.4	2514	2 F81045	hemagglutinin/hemo

ALIGNMENTS

RESULT 1

B43855

high-molecular-weight surface-exposed protein - Haemophilus influenzae

C:Species: Haemophilus influenzae

C>Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994

C:Accession: B43855

R:Barenkamp, S.J.; Leininger, E.

Infect. Immun. 60, 1302-1313, 1992

A:Title: cloning, expression, and DNA sequence analysis of genes encoding nontypeable

detella pertussis.

A:Reference number: A43855; MUID:92192797; PMID:1548058

A:Accession: B43855

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1477 <BAR>

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:89237, NCBI:89240)

Query Match

Best Local Similarity 99.2%; Score 7327; DB 2; Length 1477;

Matches 1465; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

filamentous hemagg
hemopexin-heme com
probable RTX famil
hypothetical prote
hemolysin (importe
extracellular seri
outer membrane pro
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hemagglutinin/hemo
ydek protein - Esc
outer membrane pro

April 1992 after print

Db 421 SCHYLSIESNAIVKTKWLLDPDQVTEADPLRNNTGINDFFPTGTGEASDPKKNSELK 480
Qy 481 TLTNTTISNYLKNATWNTITASRLKTVNSNIGSNHLLHSHKGGVQVQIDGDIITS 540
Db 481 TLTNTTISNYLKNATWNTITASRLKTVNSNIGSNHLLHSHKGGVQVQIDGDIITS 540
Qy 541 KGGNTIYSGGVVDVHKNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQGVTVITG 600
Db 541 KGGNTIYSGGVVDVHKNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQGVTVITG 600
Qy 601 ECKDFRANVSLGKTGKGLNIISSVNNLTHLSGTINISGNITINQTRKNTSYWQTSKD 660
Db 601 ECKDFRANVSLGKTGKGLNIISSVNNLTHLSGTINISGNITINQTRKNTSYWQTSKD 660
Qy 661 SHWVYALNLETGANFTFIKYISSNSKGLTTOYRSSAGVNFNGVNGNSFNLEKAGKVN 720
Db 661 SHWVYALNLETGANFTFIKYISSNSKGLTTOYRSSAGVNFNGVNGNSFNLEKAGKVN 720
Qy 721 KLKPNENMTSKPLPIREFLANITATGGSVFFDIYAHNSGRGAELKKNSEINSGANFTL 780
Db 721 KLKPNENMTSKPLPIREFLANITATGGSVFFDIYAHNSGRGAELKKNSEINSGANFTL 780
Qy 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNNVTLG 840
Db 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNNVTLG 840
Qy 841 QNSSSTIGNTIEKAANVTLEANNAPNOQNIIRDVTKLGSLLVNGSLTGENADIKGN 900
Db 841 QNSSSTIGNTIEKAANVTLEANNAPNOQNIIRDVTKLGSLLVNGSLTGENADIKGN 900
Qy 901 LTISESTFKGTRDTLNTGNTFNNGTAETITQGVVVKLGNTVNOGDLNITTHAKRNOR 960
Db 901 LTISESTFKGTRDTLNTGNTFNNGTAETITQGVVVKLGNTVNOGDLNITTHAKRNOR 960
Qy 961 SIIGDILINKGSLNITDSNNDABEIQIGNISOKEGNLTITSSDKINIKOITTKGIDGE 1020
Db 961 SIIGDILINKGSLNITDSNNDABEIQIGNISOKEGNLTITSSDKINIKOITTKGIDGE 1020
Qy 1021 DSSSDATSNANLTKTKELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGSGAEKTV 1080
Db 1021 DSSSDATSNANLTKTKELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGSGAEKTV 1080
Qy 1081 TFNNVSKDISADGHNVTLNLSKVKTSNGSGRESNDNDPGLTITAKNVEVKNKDIISLKT 1140
Db 1081 TFNNVSKDISADGHNVTLNLSKVKTSNGSGRESNDNDPGLTITAKNVEVKNKDIISLKT 1140
Qy 1141 VNITASEKVTTTAGSTINATNGKASITTKTDISGTISGNTVSVSATVDLTTKSGSKIEA 1200
Db 1141 VNITASEKVTTTAGSTINATNGKASITTKTDISGTISGNTVSVSATVDLTTKSGSKIEA 1200
Qy 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
Db 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
Qy 1261 GSSITSTKGQVDDLLAQNQSGIAGSINAANVTLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320
Db 1261 GSSITSTKGQVDDLLAQNQSGIAGSINAANVTLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320
Qy 1321 LNCDASGDSDEVANAVASGSGSVTAATSSSVNITGDLTVNGLNITISKGRNTVRLRGKE 1380
Db 1321 LNCDASGDSDEVANAVASGSGSVTAATSSSVNITGDLTVNGLNITISKGRNTVRLRGKE 1380
Qy 1381 IEVKYIQPGVASVEEVTEAKRVLEKVKDLSDEERETLAKLGVSARFVPPNNTITVNTQN 1440
Db 1381 IEVKYIQPGVASVEEVTEAKRVLEKVKDLSDEERETLAKLGVSARFVPPNNTITVNTQN 1440
Qy 1441 EFTTRPSQVITIEGKACFSNGGARVCTNVADDDGP 1477
Db 1441 EFTTRPSQVITIEGKACFSNGGARVCTNVADDDGP 1477

RESULT 2
A43855

high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A43855
R:Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable
detella pertussis.
A:Reference number: A43855; MUID:92192797; PMID:1548058
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1536 <BAP>
A:Cross-references: GB:U08876; GB:M4616; PIDN:AAA0527.1; PID:g475771
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBI:89235, NCBI:89239)

Query Match 66.9%; Score 4957; DB 2; Length 1536;

Best Local Similarity 67.3%; Pred. No. 1.4e-194;

Matches 1046; Conservative 146; Mismatches 265; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLALPLSAMLISLGV 60

Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLALPLSAMLISLGV 60

Qy 61 SIQSVLASLGDMVDVHGATMQVDGNKTIIRNSVDATINNKQFNIDONMVQFLOENN 120

Db 61 SIQSVLASLGDMVDVHGATMQVDGNKTIIRNSVDATINNKQFNIDONMVQFLOENN 120

Qy 121 NSAVFNRTVTSNOISQLKGLDSNGQVFLINPNGITIGKDAINTNGFTASTLIDISNENIK 180

Db 121 NSAVFNRTVTSNOISQLKGLDSNGQVFLINPNGITIGKDAINTNGFTASTLIDISNENIK 180

Qy 181 ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSGISLLAGQKIT 240

Db 181 ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSGISLLAGQKIT 240

Qy 241 ISDIINPTITYSIAAPENAEVNLGDIFAKGGNINVRATIRNOGLSADSVDKSGNIV 300

Db 241 ISDIINPTITYSIAAPENAEVNLGDIFAKGGNINVRATIRNOGLSADSVDKSGNIV 300

Qy 301 LSAGEGEAFIGVISAQNOOAKGKMITGDKVTLTKTGAVIDLSKEGGEYVLLGDERGE 360

Db 301 LSAGEGEAFIGVISAQNOOAKGKMITGDKVTLTKTGAVIDLSKEGGEYVLLGDERGE 360

Qy 361 GNGIOLAKKTSLEKGTINVSKEGGEFAIVWGDIALIDGINAAGSGDIAGTGGFVET 420

Db 361 GNGIOLAKKTSLEKGTINVSKEGGEFAIVWGDIALIDGINAAGSGDIAGTGGFVET 420

Qy 421 SGHDLFIKDNAIVDAKWLDPDNVSIINAEPLFNNTGINDFFPTGTG-EASDPKKNSEL 479

Db 421 SGHDLFIKDNAIVDAKWLDPDNVSIINAEPLFNNTGINDFFPTGTG-EASDPKKNSEL 479

Qy 480 KTTLTNTTISNYLKNATWNTITASRLKTVNSNIGSNHLLHSHKGGVQVQIDGDIITS 539

Db 479 KTTLTNTTLESLLKGTFTVNTANQRIYVNSINL-SNGSLTWSEGRSGGVEINNDIT 537

Qy 540 ----SKGNLTITYSGGVVDVHKNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQ 593

Db 538 TGDDTFRGANLTITYSGGVVDVHKNTIDQGFNLITAAVAFEGGNKARDAANAKIVAQ 590

Qy 594 GTVTITGCKDFRANVSLNGTSGKGLNIISSVNN---LTHNLSGTINISGNITINQTRK 650

Db 591 GTIT-SGNQKGFRENNVSLNGTSGKGLNIISSVNN---LTHNLSGTINISGNITINQTRK 649

Qy 651 NTS-YWQTSKSHWNVSALENTGANETFIKYISSNSKGLTTOYRSSAGVNFNGVNGNM 708

Db 650 NESGYDKFKGRTYNNLTSLNVSSEGEFNLTDSCRSDSAGTLTQPYNLINGISF---NKDT 706

Qy 709 SPNLKAGAKVNFKLPNNMNTSKPLP-IRFLANITATGGSVFFDIYAHNSG---RGAE 764

Db 707 TFNVARNVNFPIKAPIGINKYSSLNYSFNGNISVSGSGSVDFTLTLLASSNVOTPGV 766

Db 935 LTPSPKGLFRNGGWTAAANILLVANSYSGE-----TVKI-NASSNKNWITA-- 983
 Qy 956 KNRQSIITGGDIINKKG-SLNIITDSN---NDAEIOGNISQKGNLTISSDKINIKQI 1011
 Db 984 -GKDIISITAGSKTATGPNINNIENVIETNNGFTTNGITSTWLSGVNVSANGVDITSNS 1042
 Qy 1012 TKKGIDGESSDSSATSNANITIKTELK--LTEDLSISGFNKAEITAKGRDILT-IGNS 1068
 Db 1043 TGTGGI-----VLDNTNLTITVGDINTITVNSSGKGWIKSNSTLNSNKDITLVGV 1094
 Qy 1069 NDGNSGAEPKATVTFNVKRD---SKISADCHNVTLNSVKVKTSSSGGREGNSDN--DTGLT 1123
 Db 1095 AQONEG-----VIIQSSDASRNINISAGQ-NITLIGKM---GNGSGOHSILNLGNVSLT 1144
 Qy 1124 ITAKNVEVNDKITSIKTWNITASEKVTTTAGSTINATNGKASITTKTGDISGTSNGTVS 1183
 Db 1145 SSGRNIDINGSAGTGDVYFNV-----LNATAGNVSIYAEI-----KTALS 1187
 Qy 1184 VSATVLDLTKSGSKIEAKSGEANVTSATGTIGTIGTISGNTVNTANAGDLVGNCAEINAT 1243
 Db 1188 TSLNAVLSLGGNNSTKAQNG-----WLLGKAFNTQAGIGFRANSSLSVDGNILKGET 1242
 Qy 1244 EGAATLTATG-----NLTITEAGSSIT---STKQGVDLAONGSIAGSINAANVTLTNT 1294
 Db 1243 EGVGCA-TRKGDIFGANTLNIIKQSLSLGENKGAQDTAGGAGISVTSL--AKLTVNNN 1299
 Qy 1295 GTL-----TTVAGSDIKATSGTLVINAK-DAKLNQDA-SGDSTEVNAVNASGSGSVTAA 1346
 Db 1300 GSKMEGRSTSGTGINFPSNNNTLVFGDGDTLIKGSSVAGTGAAGISGVVNNSTGPMTIE 1359
 Qy 1347 TSSS-----VNITGDINTVNGLNIIISKDGRTVRLRKEIEVKYIQ--- 1387
 Db 1360 GISTDGAGVHLFSAEHRIDRINTVNTGSSTHAEGRLISGNAIVDTTLTKGSINGSGVKIDS 1419
 Qy 1388 -PGVASVEEVLEAKRVLEK--VKOLDEEREETLAKLGVSAVRFEVPNTTIT-----VN 1437
 Db 1420 LPGSSV-----TRSLVDNATLNGSSSGKGVETTSIDGIIHSSINGTTTGTGYGIDIG 1474
 Qy 1438 TQNEFTTRPSOVIISSEKACFSSGNGARVCTN 1470
 Db 1475 ENSNVGTGTEADLLILOGVATTGTGTGKLNEN 1507
 RESULT 4
 F83068
 hypothetical protein PA4625 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83068
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizioch, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: F83068
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2154 <STO>
 A:Cross-references: GB:AE004876; GB:AE004091; NID:g9950869; PIDN:AAG08013.1; GSPDB:GN001
 C:Genetic: experimental source: strain PA01
 A:Gene: PA4625

Query Match 10.1%; Score 744.5; DB 2: Length 2154;
 Best Local Similarity 23.9%; Pred. No. 1.1e-22;
 Matches 394; Conservative 231; Mismatches 585; Indels 437; Gaps 80;

Qy 87 GNKTIIRNSVD-AIINMKOFNIDONEMVQFLOENNNNSAVFNRTVSNQISQKGLDINSQ 145
 Db 60 GKTLLIDOSTQRAIINMKGFVSADEAVFRNPQGVTSSTLNRVTAQGVSIAGRISAPQ 119

Qy 146 VFILNPNGITIGKDAIINTNGFTASTLIDISNENIKARNFTFEQTKDALAEIVNHHGLTV 205
 Db 120 VIYNSNVVFSGSAKVDVGLSLTTITANISDEHFQKGLIFDQPNPD-ARIIVNDGSISV 178
 Qy 206 GKDGSVNLIGKVKNEGIVSVNGGSSISLAGQKITI---SDIINPTIITYSAAPEA- 260
 Db 179 AEKGLAAFAVPSVANNGVINARLGTVMAAGNAATIDLVGDGLVSTAVTDPVTRKPDQAQ 238
 Qy 261 ---VNLGDIFAKGNNINRA---ATIRNOGKLASDSVKD---KSGNIVLSAKKEAEI 310
 Db 239 ALVNSGAIQADGQSVLITAEQASRVVDNAVNLGVLARGTEVRGSSVALVSKSDIQI 298
 Qy 311 GGVISAOQOAKGKLMITGDKVTLKTGAVIDLSG-KEGETYLLGDERGEKNGIQLAK 369
 Db 299 AGKIDVSGPK-NGCDVLVSGQVALASTASIDARTAGQGSVRIGGDFQGRGE--LPRAK 355
 Qy 370 KTSLEKGSITNVS--CKEKGGAIVWGD--IALDGINAOGSDIAKTGGFVETSGH-DL 425
 Db 356 NATLAKGASIDVSATCKGNGGLAVVWSDGNTKMDGRILARG-GAOGNGGLVETSGKVL 414
 Qy 426 FIKDNAIV-----DAKEWLLDFDNVSI NAEDPLFNNTGIDNDEFPTGTGEASDPKKNSE 478
 Db 415 SIADSAVSVAAPIYNGGTWLLDPTTLRIVA-----SGTSGSVGGANGASGSDATVNAS 468
 Qy 479 LKT-TLTNTTISNYLKNAMTMITASRLKTVNSSI---NIGSNH-LILHSGQREGGVQ 533
 Db 469 VVTGALAGGKVT-----LSASDRLSVEAPLITSNLGASRGLELIATGP-AGAVD 517
 Qy 534 IDGDIITSK-----GGNLTIYSGGVVDVHKNTILDQGLF---NITAAVSAFEGNNKA 582
 Db 518 ISAPILFRNGSLAIRAGGNINFLSGGTPQTSIGIVDLGSGTLAMQTSIAGKISQOAGT--A 575
 Qy 583 RDAAN-----AKIVAQGVITITGEGKDFRANNVSLNG-----T 615
 Db 576 LIAANLAGRASDILASWDNYAGNALQFNGLTKVRSNATGVTTISGTVDFPFTNQSM 635
 Qy 616 GKGLNIISSVNN--LTHNLSGTINISGNTTINQTRKNITSY-----W 655
 Db 636 GTAQNIIVSSVGTIRLEANSVGI---TGNITL--TAGNSEFDRLVFTALPYRRVSGSASF 690
 Qy 656 QTSHDHNVNVSALNLE-TGANFTFK-----YISSNSKGLTT---OYRSSAGV-NFN 702
 Db 691 PTNDSYDLVTLNLYQVNGSNVTATPNGAPSGFTVAAGNGSVTTTGNWGTSMWKVGF 750
 Qy 703 GYNG-----NMSFNLKEGAKVNFKLPNNEN----- 728
 Db 751 GVIGVTDELQYDVGTLTBEILIFGLGKTSRVDTRLDLFMREGAFNSFAERAQVEMFKTT 810
 Qy 729 -----NTSKPLPIREFAN-----ITATGGSVFFDIYANHSRG---AELKMS 768
 Db 811 TTAGDILSRQQTATILTANDATRVYGDVNPVTLTATWSGINADAYVNSQFNDLYQATAST 870
 Qy 769 EINISGANFTLNSHVGRDAPK-----INKDLTINATNSFSLRQTKDDFYDGVARNAI 823
 Db 871 ATQASNVQYAITGANGSEYFSQRYQLVRQDRLVTTPAQLIVSADAKTKVYGDADPTL 930
 Qy 824 NSTYNIS-----ILGNV-TLGGNSSSSTITGNITTEK-----AANVTL----- 861
 Db 931 --TYQVSLKNSDTAAGVLSGNLGRVAGEN-----VGNVGLQGGGLGLANTANYTLUSYGN 983
 Qy 862 EANNAPNOON-IRDRVIKL-----GSLLVNGSLS-LTGENADIK 898
 Db 984 DLURITPAQLNVIAADAKTKVYGDLPALTYQVYSGLKRGTAGAVLNGGSLSRVAGENVGV 1043
 Qy 899 GNLTISESATFKGTRDTINITNGFTNNGTAINITQGVKL-----GNVTNDGDLNITT 953
 Db 1044 G---INQGGGLGVSSNYTLVQGN-----NLTTKALLNVIAADAKTKVYGDADPALTY 1093
 Qy 954 HAKRNORSITIGDIIN-----KKGSIINITSNNDAEIOIGNISQK 996
 Db 1094 QVSGKLNKGDGTAGAVLNGGSLSRVAGENVGVYINGOGGLLSANYDLSYQ--GN----- 1145
 Qy 997 NLTISSDKINITKQITIKKIDGEDSSSDAISNANLTKTELK-----LT----- 1042

Db 1146 NLITKALLNVIADAKTKVGDADPS-----LTYQVSLKNGDGTAGSILTGGLNRA 1196
Qy 1043 --EDLSISGFNKAETIAKQGR-DLTIGNSDG-----NSGAFAKTVTFNNVKDSKI--- 1090
Db 1197 AGENGVYGINQDGLALNSGNYDLSVQGNLITITKALLNVIADAKTKVYCDADPSLTYOV 1256
Qy 1091 ----SADGHNVTLN--SKVKTSSNGGREGSNDNDTGLITAKNVEVKNKDTISLKTVNIT 1144
Db 1257 SGLKNGDTAGVNLNGGLVRVSGENVGYAIOQGGUGL-----VSGNTDL-AYOGNLIT 1309
Qy 1145 ASEKVTITAGSTINATNGKA--SIT-----TKTGDISGTI-----SGNTVSV----- 1184
Db 1310 ITRKALLNVIADAKTKVGDADPSLTYQVSLKNGDSAGSILTGGLNRAAGENVGYGINQ 1369
Qy 1185 -----SATVDLT-----TKSGSKI--EAKS-----GEANVTSAITGIGTISGNTVNV 1225
Db 1370 GDLALNSGNYDLSVQGNLITITKALLNVIADAKTKVYGDAD-PSLTYQVSLKNGDTAGA 1428
Qy 1226 TANAGDLTVNGAEIN--ATEGAATLTATGNTLTTEAGSSITSTKQVDLLA----- 1275
Db 1429 VLNGGLVRVSGENVGYAIOQGGUGLGLVSGNYDLAYOGNLITITKALLNVIADAKTKVY 1488
Qy 1276 -----QNGSIAGSINAAVNTLTGTLTVAGSDIKATSGTLVINAKDAKLN 1322
Db 1489 DADPSLTYQVSLKNGDTAGV-----LN-GGSLSRVAGENV-----GVYGINQDGLALN 1537
Qy 1323 GDASGDSDEVNAVNASGSGSVTAATSSSVNITGLNT-----VNGLN----- 1364
Db 1538 ---SGN-----YDLSVQGNLITITKALLNVIADAKTKVYGDADPSLTYQVSLKNGDRA 1588
Qy 1365 --IISKDGRNVTLRGKEIEVKYIQPG 1389
Db 1589 GAVL--NGGGLVRVSGENVGYAIOQG 1613

RESULT 5
A:3080
hypothetical protein PA4541 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1417 <STO>
A:Cross-references: GB:AE004867; GB:AE004091; NID:9950769; PIDN:AG07929.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
C:Gene: PA4541

Query Match 8.6%; Score 636.5; DB 2; Length 1417;
Best Local Similarity 21.9%; Pred. No. 1.6e-18;
Matches 352; Conservative 238; Mismatches 523; Indels 495; Gaps 72;

Qy 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHRLAKPLSAMLISLIGVT 60
Db 1 MNKSYTLWNQ-----ATGCWNVASEGTTRRSK-SGRCKALVVGASILLGL-FC 47

Qy 61 SIPOSVLASGLOGMDVYHGTATMQ--VDGNKTI--RNSVDALINWKFNDONEMVQFLO 117
Db 48 QAPAFALPS-----GATVSGDAGFGTSTDGRHVMYDQOSHKLITWNFEFSVRADERVSFHQ 104

Qy 118 ENNSAVFNRVTSNQISOLQILDNSQGVFLINPNTIGKDALINTNGTFASTLDSNE 177
Db 105 PGQDAVALNRVIGRNSDIOGRIDANGKVLVNPNGVYFGKSAQVNVGGLVASTLDIADR 164

Qy 178 NIKARNTFEOTKALAEIVNHGLITVKGDSVNLIGCKYKVBEGVISVNGSSISLLAQ 237
Db 165 DFLAGNYQFSGDSG---ATVSNAGSLQASEGSSIALLGARVSDNGLTQAOLGDVALGAG 221
Qy 238 KITSDIINPTITYSIAAPENAEVNLGDFAKGGINVRA-----ATIRNOGKLSADSV 291
Db 222 GINLN-----FDGDGLLNLOVDKGSVDALAHNGGLIRAD-- 255
Qy 292 SKDKSGNIVLSAKBEAEIGVISAON-----OAKGGKLMITGDKVTILTKTCAVDLS 344
Db 256 ---GGQVLSARSADSLTKTVVNNQGLTEARTLSAEGRIVLDDGEGOTVVRAGKODAS 311
Qy 345 KEGGETYLGDEREGKNGIOLAK---KYSLEKSGTINVSQKKEKGGFALVWGDIALID 400
Db 312 AIGGN---GGVLVNOGAN-VEIORTAQVDTHADGAT-----GTWRILSHEVSVA 359
Qy 401 -GNINAOSGDIKTGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSIANAEDPLFNNTGI 459
Db 360 VQANAAGDG-----SGOVHVAQGPAGANAS---DSNGVTIVOOOP----- 397
Qy 460 NDEFTGTGEASDPKKNSLKTTLTNTTISNYLKNATMNITASRKLTVNSISINSGNSH 519
Db 398 ---AVDLAAGANGTSVOSQSGANGISVVQSQ-----NSPNIGSGAN 442
Qy 520 LILHSGGORGGV-----QIDGDTISKGNLTITVSGMWVDVHKNTITLDOGFNL 567
Db 443 GLSVVOSONGANIGAGASGISVVOSONSPNIGSGVNGVTV-----VQSONGANIGSGASG 497
Qy 568 ITAASVAFEGGNKAKARDANAIVAGTIVITGEKDFRANNVSLNGTGKGLNIISSVNN 627
Db 498 ITV--VQSONGANIGSGASGISVVOSONSPNIGSG-----VNGVTIVQSONGANIGPVSG 551
Qy 628 L-----THNLSGTINISGNITINOTTRKNTSYWQTSHD---SHWVNSALMLETGANFT 677
Db 552 IDVQOTQLPNLSPGANGSSIVQV-----OTLPDIAADAGNVHVVOVQGGNKV 600
Qy 678 FIKYISSNSKGLTQYRSSAGVNFNGVNGMNSFNLEKAGKVNFKLKPENMNTSKPLPIR 737
Db 601 F-----GNSATNYSR---TVOARSNENVSG----- 624
Qy 738 FLANITATGGSVFFDIYANHSGRAELKMEINISNGANFTLNHSH----- 784
Db 625 -LANPSSAGKGT---LHADTLARLSTSNVEVATRG-----NAHVAPLSDWSGNGLT 675
Qy 785 ---RGDDAFKINKDLTINATNFSLSRQTKDDFYDGYARNAINSTYINISILGGNVTLLG 840
Db 676 LTAERGD--LRINGALTAQENASLTNA-----GORPLRIDDLSLTGOCARVEF-- 724
Qy 841 QNSSSITGNITIEKAANVTLEANNAPNOONTRDRI-----KLGSLLVNGSL 888
Db 725 -NSDK---GYALAEGRITLTSKKNAGFRANGDYSVIQLQDLRGIDRLDLSGYSVLGN- 778
Qy 889 SLTGENADIKNLTISESATF-----KGTEDTLNI-TGNFT-----NNGT----- 928
Db 779 RIAGGNSSF---LSIGNASAFGCTFDGLNTIDNLAVYCTGAYSGLFSVNRGTLRLNLE 835
Qy 929 -----AEINITQG---VVKLGNV--TNDGDLNITHAKRNORSIIGDIIINKKSL 974
Db 836 RISADGAQATHYVQVGSAAVNLGRIDNVNASDIRIAAASKLNS---LGGVLVALNLSGI 892
Qy 975 NITDSN-----NDAETIQGG---NISQKEGNTLITSSDKINITKQITTKKIDGEDSSD 1025
Db 893 DNASAGTLVGNRHTYALGGLAAENISTARGVASISNSR----- 931
Qy 1026 ATSNANLTIKRELKLTEDLSISGFKAEITAKDGRDLTIGNSNDQ---NSCAEAKTVTF 1082
Db 932 -----ADFAISGOLK-DHASYGAGGLVGNRGLIRSSGQ----- 967
Qy 1083 NNVKDSKISADGHNVTLNSKVKITSSNGREGSNDNDT-----GLTITAKNVEVKNKDT 1136
Db 968 ---GTLSSHGMMGLGVYSSAGGLADVASVDVSGNQRGLYGLIGLGNVSGI- 1021
Qy 1137 SLKTVNITASEKVTTT-----AGSTINATNGKASITTKTGDIS-----GTI--- 1177

```

Db 1022 ----AHATASGKVRGTDABALGGLIGLNNAANNAS-----AHGDSVSLQAGRYLGLIGHN 1074
QY 1178 -SGNTVSATVDLTPKSGSKTEA-----KSGEANYT-----SATGTIGT 1217
Db 1075 OAGNLAVSTSGNLS--GGSLQAGGLIGLNANASLVNASAKGNVATRGAEAVGGLGEN 1132
QY 1218 ISGNTVNTANAGDLTVGNCAGINATEGA-----ATLTAT----- 1252
Db 1133 LVGSVINGSA-SGEVDSGKTLGGLIGLNHNSLNKASGVWAGANSVDVGLIGHNR 1191
QY 1253 -GNTLTTEAGSITSTKGOVDLLANGSTAGSINAANYT-LNTTGLTT-----VAG 1302
Db 1192 GGNHSTLAASGVNTGKG-----SRVGLGVYDAASTNVNSASGVNSASGRATGGLIG 1246
QY 1303 SDIKATSGFLVINAADAKLNGDAS-----GDSTEVNNAVNASGS----- 1340
Db 1247 SCLR--GSLMLASSHGIVNDKTNHNLGGLVGRGENTSIKASAKASGVSGAGIRAGLV 1303
QY 1341 -----GSVTAATSSVN--ITGDLNLTNGLNLIISKDG 1370
Db 1304 GSLEGWQALILGASAGDVTAGYSDYIGGLVGFSTATISGASASCKVG 1351

RESULT 6
AD0123
probable autotransporter protein yaph [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0123
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Jl. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0123
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3705 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89847.1; PID:g15979073; GSPDB:GN00175
C:Genetics:
A:Gene: yaph

Query Match 8.1%; Score 602; DB 2: Length 3705;
Best Local Similarity 21.7%; Pred. No. 1.4e-16;
Matches 413; Conservative 260; Mismatches 695; Indels 534; Gaps 91;

QY 14 NALVAVSELA--RCDDHSTEGS--EKPARMKYRHLAKPLSAMLISLGVTSIPQSVLA 68
Db 972 SGVTVTDLALNSGVNTIGSSGNDNDPLATNVSILODRAGSTLELINATNVGDDA 1031
QY 69 SGLOGMOVHCTA-TMQVDGKNTIIRNSVDALINWKOFNDQNMVQFLOENNNNAVFN 127
Db 1032 LDL-----LVNGTALTSGTGVSALIQGGSTVA-----NATHYGLASSNSGDSGLYVN 1082
QY 128 VTSNQISOLKILSDNSGVFLNPINGITIGKDAIIN-----T 164
Db 1083 YT--LSALELLADGADALLATSEGLTANR--VLNAELFVGGLVVDQNGALTLANGS 1137
QY 165 NGFTASTLDISNENIKARNFTFEQTK-----DKALAEIVNHGLITVKGKG 209
Db 1138 NRYEGTTVTAGELILGALGAFGQTSLLDIASGASANINGYSQTVGAVTVNGVTLGSG 1197
QY 210 SVNLIGKVKRNEGVISVNGGSIISLAGOKITISDIINPTITYSIAAPEAVNLGDI 269
Db 1198 V--LTSGLTNGGLDGLTGALNLTAGASTVAGGLTGAGTLN-----N 1240
QY 270 GGNINVRRAATIRNOGKLS-AD--SVSKDKSGNIVLSAKE--GEAEITGGVISAQNOQAKG 324
Db 1241 GGNLSVSAANSGLSGOHIADVASVTLTDIGTLCTSAVEVLGTNLNNGANAATN----- 1295
QY 325 KLMTGDKVTLKTCVADILDSKE--GGEYTLGQDER-----GEGK----- 362

```

```

Db 1296 ----VLSGDDG-TINTNAAVTLSSGNSFSGAHQIGTGDGLTVGQASNLGASSATVNLGLTSH 1352
QY 363 ---NGI--OLAKTSLKSGSTINVS-----KEGGFAIVMGDIALI-----DG 401
Db 1353 LILNGVSESIANVSUGVAGSTVDIIGGADTALTANNSGFL---QOYALAGNSKLTVA 1409
QY 402 NINAQ-----GSGDIAKTGTFVETSGHD-----LFKDNAIVD-----AKEWLLDF 442
Db 1410 NLGASSSVALLAGADGTLTSLGFGNCTFGNSVTSVGLQVTDAAEVLTSNGSVNAV 1469
QY 443 DNVSINAED-PLFN-----NTGIND-----EPTGTGTEASDPKKNSELKTLTNTT 487
Db 1470 ADAFLNLDLDALENHVLNGLLNNAKNDASTAFDFGTVGGAFSGIVN-----LTNTT 1523
QY 488 ISNVLLKNAWNTMTASPKLVNSINIGSNHLLISKGGGQGVQIDG----- 536
Db 1524 FALSADNAAL-ARATLKLSDSVTVTVGATDR-TLHGLDLNGTGLTFDGSPPSOQANGVV 1581
QY 537 ---DITSKGNLTIY-SGGWVDVH-----KNITL---DQGFNLTAASVAFEGGNKARDA 585
Db 1582 TVTDLALNSGTISITGAGNWNENHPVTPPNVSLLEQDRGDILLELELNAANVTGN-----A 1636
QY 586 ANAKIVAQGTVTITG-----EGKDFRANV-----SLNGTG----- 616
Db 1637 NNLDDLVDGTAITSGTQGVESAIOQGGSTVANAHHNYGLTSSNGNGSGGLYVNTLSALE 1696
QY 617 ---KGLN--IISSVNNLTHNL-----SGTINISGNITINQTRKNKTSYMWOTSHD 1668
Db 1697 LLANGANALLLATESGLTANRVNLNAELFVGGLVVDQAQNGALTLANGNNRYEGT 1756
QY 669 NLETGANFT-----IKYISSNSKGLTQVRSASGVNFNG-----VNGNM 708
Db 1757 ELILGANGAFQTSLLNIAASGASANINGVQTVGAVTNSGAVTLGNGVLTSC 1816
QY 709 -----SENLEKCAKYNF-----KLKPN-----ENMNTSKPLRPLANITATG 746
Db 1817 LDLTGGLNLAAGSSSTVAGGLTGAGTLNGLDGLAVSATNSGLSQTHIADVASV 1876
QY 747 GGSVFFDIYANHSGRCAELKMEINISNGANFTLNHVRGDDAFKINKOLTINATNS 806
Db 1877 TGTL-----GTSAVEVLGTNLN-NGANAAMTVLSSGGVINTNAAVT 1926
QY 807 LRQ--TKDDFYDGYARN--AINSTYINISILGGNVTLGGONSS-----SSITGN-ITIEKA 856
Db 1927 AHQIGTGDGLTVGQASNLGASSATVNLGLTSHLILNGVSESIANVLSVAGSTVDI 1986
QY 857 ANVTLEANNAPNOQNIIRDRVILKLVNGSLSLTGENADIKGNLTISESATFKGTRDT 916
Db 1987 ADTALTANNNG-----FLGQYALAGNSKLTVA 2032
QY 917 LNTGFTNNGTAEINIT-QGVVVKLGNTVNDGDLNITT-----HAKRNORSI 962
Db 2033 LSLSG---FNGTFGNSVTSGVLQ---VTDDAEVTLTSSNGVGVNTVKVDIADATL 2086
QY 963 IGGD-IINKKGSINITDSNDAEIQIG-----GNISQKECNLTISDDKINIKQITI 1013
Db 2087 ALFDHVLTCNGTLNVAKNLATTAFDFGTVGAFSGIVNLTNTTFALSADNAALARA 2146
QY 1014 KKGID-----CED-----SSDATSNANLTKTKELKITE-DLSISGFN 1051
Db 2147 KLSDDSVTVTGTTDRILHGLDLNGGTLTFDGSPPSOQANGVVTVTDLALNSGTISIT 2206
QY 1052 KAE-----ITAKDGRDLTI-----GNSNDGNSGABAKTVTFNNVKDSKIS 1091
Db 2207 WNEHEPVTTPPNVSLLEQDRGDILLIOLIDADNVNTGNANDLELMINGTISAGQGV 2266
QY 1092 ADGHNVTLSNKVKTSSNGGR-----ESNSD-----NDTGLTITAKNVEVN 1132
Db 2267 OGGYTVANATHNYGMTSNGGSGLYVNYTLTSALELLADGANALLLATESGLT---ANREL 2323
QY 1133 KDITSILKTVNITASE-----KVTTTAGSTINATNGKASITTTKTGDSGTIS 1178

```

Db 2324 AELSGVGLVDAQNCALTLANGNNRYEGTIVTAGELILGANGAFGOTSLNIASSA-S 2382
QY 1179 GNTVSVATVDLTGSGKIEAKSGE-----ANVT-----SATGTIGGTI 1218
Db 2383 ANINGRYQVAVTGTGVLGNGELISTDPLINTGMINVTDGLINLENGGASSISGL 2442
QY 1219 SGNVTVNANAGLIV-----CNGAEINATECAATLTATGNTL--TTEAGSSITSTKQV 1271
Db 2443 TNGI-LNTKGGDFTSIDNGLAGTNIISDCAVTLGNGGTIGTIGLNGSSVIDVLDGL 2501
QY 1272 DLLAONGSTA-----GSIN-AANTVLA-----TTGTLTTVAGSDIKATSGT- 1311
Db 2502 NLVADN-SLANVYSGDGTINTATVTLSGNSSFSGAHQIGTNGELTVQASNLGASSATV 2560
QY 1312 -----LVINAKDAK-----LNGDA-----SGDSTEVAVNA-----SGSGVT 1344
Db 2561 NLGTLTSHLILNGVSESIANVLSGVAGSTVDIIGADTALTANNSGFLGQVALAGNSKLT 2620
QY 1345 AAT-----SSSVNI--TGDLNVTNGLN-----IISKDG-RNT 1373
Db 2621 VASTNNLGASSVALAGTGDITLSLGFNGTFCNSVTGCVLOVTDADAEVTLTSSNGVSA 2680
QY 1374 VRLRKEIEVKYIOPGVASVEEVEIAKRYLEKVKDLSDEERETLAKLGVASVRFVEPNNT 1433
Db 2681 VTIDIADATLNL--DDIALFNIALTGNGLLNVAKNDASTAFDFGATVGGAEFTGTYNLANS 2738
QY 1434 ITVNTQNETTRPSSOVIISSEKACFSSGNGARVCTNVADGG 1475
Db 2739 TFDLSGNTTVLAQATLKLSSGNTL-SVGNQVQNTGLAMNG 2779

RESULT 7
T31105
hypothetical protein 2 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31105
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J Bacteriol 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference numbers: 220984; PMID:99030326; PMID:9811662
A:Accession: T31105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4919 <NAR>
A:Cross-references: EMBL:AF057696; NID:g9329021; PID:g9329023; PIDN:AAC79761.1
A:Genetics:
A:Gene: lspA2

Query Match 8.0%; Score 593.5; DB 2: Length 4919;
Best Local Similarity 23.0%; Pred. No. 4.4e-16;
Matches 385; Conservative 226; Mismatches 604; Indels 459; Gaps 83;

QY 2 NKIYRLKFKRLNALVAVSELARGCDHST-----EKGSEKPMKVR-----H 44
Db 3 NKRYKLIFSCKVKKLPVPAENIKSASGNSGSSSKIAEDQEEPDLSLACSLSPSSSIH 62
QY 45 LAL-----KPLSALLSLGVTSTIP-QSVLASGLQ-MDVVHGTATQVD----- 86
Db 63 LGLNHSPLKVFVKGNLSVLLSL---MPAMQVWADSSNAIVDHSHGAKQTAVDERDPK 118
QY 87 -GNKTIIRNSVDIAIN-----WKQFNI-----DQNMVQVLOENN 120
Db 119 NGREKV-----VINIKPDEQISDNHFSKFNIPNSAVFNNSIREGNSQLVGLLGENK 172
QY 121 N-----SAVENRVTNSQISOLGILDSNGQ---VFLINPGLITICKDAITINGFTAST 171
Db 173 NLGSOAKTIFNOVTDQESKISGLLEVGEKADLFIINPGLVLCVKTINTDREAVST 232
QY 172 LDISNENIKARNFTFEOTKDKALAEIVNHLITVKGDK-----SVNLIGGKVKNEGYI 224
Db 233 SEVVEPHIKQLN-----VORGKVIIGDKGVATNGLSHFDDVVAKNIEQOGKV 278

QY 225 SVNGGS-----ISLLAGOKITISDIINPTITYSTAAPEAVNLGDIKFAKGNINVRA 277
Db 279 SIEGDSKPAKLANVTFAAG-----NLTYD-----VNRDVT-----NRNT-- 312
QY 278 ATIRNOCKLSADSVDKSGKNIVLSAKEGAEIGGVISAONOQAKGKMLITGDKVTLTK 337
Db 313 ----NPKKPIVDNRKD--NIAIS--GE-----SAGSMYGRNIKFIVT----- 347
QY 338 GAVIDLKSGEGGYLGGDERGCKN--GIOLAK-----KTSLEK--GSTINVS 383
Db 348 -----DKGAGVNHQGVIFAEDDINILTDGNSRLNKVADYVRVVG 388
QY 384 KEGGPAIVGWDIALI--DGINAOGSGDIAKTGTFVETSGHDLFIKDNAIVDAKELLDF 442
Db 389 K-----DIELANNGQIHADQOQLILNATGHVKLNDSSVSNLIGSALN--LTL 436
QY 443 DNVSINAEPLF---NNTGINDPPTGTGEASDPKKSEL-----KTL--TWTTI 488
Db 437 ENATVSANLFRVTNDTKLNLKSVARADLOSGLNLDKASVLAHLKLTINISDVSL 496
QY 489 SNYLK-NAWTNITPASRLTVNSS-----INIGSNHLILHSGQKRGG-----V 532
Db 497 NNQKLSANNLKIKKVRDLNINNSLSANNLTLATSNITLKNKSKFTAGMTLNTVNV 556
QY 533 QIDGDIKSGGNLTYSGGWVDVHKNTLDQGF-----LNIT-----AAS 572
Db 557 TLNNDSELAANLTL-----NVTKNVTLNDASKLSANKLDLNTVNTLNSKSTLSAGE 610
QY 573 VAFEG-----GNKARDAANAKTIVAOQTVITGEK-----DFRANNVSLNG----TG 616
Db 611 LTFKKVKNVTLNNDSELAANLNLNASHNVTLNKSKLSAQAKADIKAVNLTLNDTTLTA 670
QY 617 KGLNIISVNNLTHNLGTI--NISGNIITNQTTRK-----NTSYWOTSHDSHW--NV 665
Db 671 KNLDI-----NSTTITNNGTIAGIFANITTEKLNKKEKALILAEQNLNF--TVNGSHYENK 724
QY 666 SALNLETGANFTFK--YISSNSKGLTYQVRSAGVNFNGVN-----GNMSFNLKE 714
Db 725 GDIVSKDKATVTFKNSDFTSNGSKLVNAQNLKVVNNVNETISQDDITLIGNVTLN--AS 783
QY 715 GAKVNFKLKPNENMTSKPLRIFLANITATGGGVDFDIYANHSGR-----GAELKWEI 770
Db 784 GTFTN-----SGNLTVTKLDVGDIQNTFNKGLNLTGVEDLHKSKTKITNDGKLISKNL 838
QY 771 NISGANFTLNSHYRCDAPKI-----NKDLTINATNS--NPSLQTKDDFDYDGYARN 821
Db 839 NISSADFINNGTILGIEALKIATKGNFTNKEKAILASNLSDISVABGKKTENN----- 894
QY 822 AINSTYNISILGGVNTLGGQNSSSITGNITIEKAAVNTLEANNAPNOQNIHDRVILGS 881
Db 895 TIESGKNLNTNWTGAFLNVYDNATIRSGVLNITSTGVN--SNN-----GT 937
QY 882 LLVNGSLTGNADIKNLITISESATPKGTRDTLNTGFTN--NGTAEIN-ITQGVV 938
Db 938 LISNERLNTS-----AANFTNESGTVMSNGLLIIIA 970
QY 939 KLGNTVNDG-----DLNITTHAKR--NORSIIGG-DIINKKGSUNITDSN--NDAETQ 986
Db 971 KOGNITNKNLTAASROQLNLTAVADNITNDSNIKNIAVLHLSGLNLSLASKQDVYNLGEY 1030
QY 987 IGGNISQKEGNTLISSDKINITKQITIKK-----GIDGEDSSSDATSAN 1031
Db 1031 AGNNISVRAHQL---KNDVKLMGDITTKTKEGQASYKLYQASNGHGFNGDSS--GYSBGD 1086
QY 1032 LPIKTKELKLTEDLSISGFENKAEITAKDGRDLTIGNSDNGSGAEAKTVTFNNVSKDIS 1091
Db 1087 LNIKGFADLDNKLTVORIGK--IYA--GRDLTFNKSNAAGK-----SEII 1128
QY 1092 ADGHNVTNLSKYKTSSSNGGRESNDSNDTGLTITAKNVEVNVKNDITSLKTVNITASEKYVT 1151
Db 1129 NRG---TINVKNKLSVDSVFNENMQSOKVDLYTKIFEAKSDIE--LTFKNGTHPVVL 1183
QY 1152 TAGSTINATNGKASITTKTGDISIGTISGTVSVSATVDLTT--KSGSKIEAKSGEANVTSA 1210

Db 1184 NFKSNNEKRYRSENTEKFNKFSIGDLINELASDAPAEIAEYSGS-----SSNVINPVSY 1239
 QY 1211 TGTIGGTTGNTVNVANAGDLTVGNG--ABINATEGAATLTATGNTLTTEAGSSITSTK 1268
 Db 1240 LAALGANNSNPHYLNTALKHILGNGWQDLKQENIKLVKQWEDFKDKGAS----- 1294
 QY 1269 GOVDLAQNGSIAGSINAANTVLTGTLTVAGSDIKATSGTLVINAKADKLN-GDASG 1327
 Db 1295 KMLDLPNTDEKAKI-PAGIIRNGNDTISDVESEDEKKYYSKQ-NGEAKNDTGTDSY 1352
 QY 1328 DSTEVN-----AVNAGSGSVTAATSSVNITGDLNTVNGLINIISKDGRVTLRKG--E 1380
 Db 1353 DSTKASEKRYKVENVDHKENI--DEHKLNIGKHEITVPGVSPENLNKKNWDHQPDKLGE 1409
 QY 1381 IEVKYIQPGVASVEEVEIAKVLKVKD--LSDEERETLAKLGVSAVRFVEPN 1432
 Db 1410 IDKSIISELLA--OPVYTEKSAARDSPRVQNQDKALONLYKTRLSYINQNN 1460

RESULT 8
 A35140
 hemolysin A precursor - proteus mirabilis
 C:Species: proteus mirabilis
 C:Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 08-Oct-1999
 C:Accession: A35140
 R:Up hoff, T.S.; Welch, R.A.
 J. Bacteriol. 172, 1206-1216, 1990
 A:Title: Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin gene
 A:Reference number: A35140; MUID:90170827; PMID:2407716
 A:Accession: A35140
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1577 <UPH>
 A:Cross references: GB:M30186; NID:gl50888; PIDN:AAA25657.1; PID:gl50890

Query Match 7.5%; Score 555.5; DB 2; Length 1577;
 Best Local Similarity 20.7%; Pred. No. 3.7e-15;
 Matches 361; Conservative 257; Mismatches 606; Indels 523; Gaps 79;

QY 40 MKVRHLAKPLSLGVTSTIPQSVLASGL-----QGMDDV---HGTATQV---DGN 88
 Db 1 MKNFKLSPSGRLAASLAIFVSLNAYGNIYDPDAGHQGPDVSAGVNGQVIVNITVNN 60
 QY 89 KTIIRNSVDAILNKKQNI-----DQNMVQFLOENN-----SAVENRYT 129
 Db 61 EGISHN-----QYQDFNVGPKGAVFNNALEAGQSQLAGHLNANSNLNGQAASLIINEVV 114
 QY 130 SNOISQLKG---ILDSNGQVFLINPNCITIGKDAINTNGFTASTLDIENIKARNFTF 186
 Db 115 SRNPSFLGQOEVEFGIAAEVVLNPNICITCDGCGFINT-----SRSSLVVGNPFLF 164
 QY 187 EOTKDALAEIVNHLITVCKD-----GSVNLGGKVKMGVSVNGSGSISLLAGOKIPI 241
 Db 165 ENGOLKGYSTLNNLTSLGKNGLNTTGLDLIAPRIDSRG--KITAABEISAFITGON-TF 221
 QY 242 S---DIINPTITYS-----IAAPENAEVNLGDIPIAKGNNVNVRAATR 281
 Db 222 SQHFDILSSOKPVSALDSYFFGWSQSGRIRINTACSGVKLAGKTAENDLSRADNIQ 281
 QY 282 NQKLSADSVSKDKSGNI-----VLSAKEGEAEIGGVISAQNOQAKGK 325
 Db 282 TDSQRYDSYDKDGSQENYQNYRGITVNNSSGSQTTLTKELKGNITLVASSHNOIKASD 341
 QY 326 LMITGKVTLTCAVIDLGSKEGGTYLGGDERG--EGKNGIQAKKTSLEKSGSTINYSG 383
 Db 342 LM--GGDITLQ--GADLTIDGLOQKQETIDNRWFYSWYDVYKKEQIQIGSQLD--A 396
 QY 384 KEKGGFAIVWGDIALIDGNINA-----QCSGDIAKTG----- 415
 Db 397 KNNATLTATKGDTLDAAKINAGNNLAINANKDIIHNLGVEKRSSENGKRNHTSRLES 456
 QY 416 -----GFVETSGHDLPIKIDNAIVDAKEWLLDFDNYSI 447

Db 457 GSWNSHOTETTLKASELTAGKDLGLDAQGSITAOQAKLHANENVLYNAK-----DNINL 510
 QY 448 NAEDPLFNNTGIDNEFTGTGEASDPKKNSEKLTTLTNTTISNYLKNAMTMTASRKLIT 507
 Db 511 NVQ-KTNNDKVTVDNIVMVGCGGGKNNNNQOQVSHAT-----QLTADGQLL 558
 QY 508 VNSINIGNSHILHLSKQORG-----GGVQIDGDIITSKGNLTIIYSGWVDV---H 556
 Db 559 LAADNNVNITGSQV---KNGQAFVKTTCQDVVIDNALSETISKIDERTGTAFNITKSSH 615
 QY 557 KNITLDQGLFNITAAVAEFGGNKARDAANAKIIVAGQTVTITGECCKDFRANNVSLNGTG 616
 Db 616 KNET-----NK-OTSGSELISDAOLTIVVS-----GNDVNVIG-- 647
 QY 617 KGLNISSVNNIETHNLSGTINI-----SGNITINQ 646
 Db 648 ---SLIKSADKLGIIHSLGDIIVKSAQVTKIDDEKTSLAITGHAKVEDEKQYSAGFHITH 704
 QY 647 TTRKNTSYWQTSHDSHNVVSALNLETGAN--ETF-----IKYISSNS 686
 Db 705 TTNKNTS--TETEQAANSIISGANVDLOANKQVTFAGSDLKTTAGNASITGDNVAFVSTEN 762
 QY 687 KGLT--TOYRSAGVNF--NGVNGNMSFNLKEGAKVNF---KLKPNENMNTSKPLPIRELA 740
 Db 763 KRQTDNTDTTISGGFSYTGVD-----KVGSKADFOYDKQHTQTEVTNKRGSQTEVAG 815
 QY 741 NATATGGGVSFFDIYANH--SGRGAELKKEISEINSGANFTLNSHVYRGDGAFAKINKDLTIN 799
 Db 816 DLTITANKDLLHEGASHHVEGRYQE---SGENIQHLA--VNDSEYKSDLSLVNGDVGVN 870
 QY 800 ATNSNPSLRQTKDDFYDGY-----ARNAINSTYINISIL-----GGNYTL 838
 Db 871 LDYSGVT-KPVKKAIEDGVNTTKPGNNTDLTKKVTARDAIANLANLSNLETPNVVEVGI 929
 QY 839 GQNSSSSITGNIITEKAAAN---VLEANNAPNOQNIIDRVIKLG-SLLVN----- 885
 Db 930 KGGGSOQSDTDSQAVSTSNAGKIDIDSNNKLDGTHYQSTOEGISLTANTHSEATLD 989
 QY 886 -----GSLSLTGEN--ADIKMLTISESATPKGKTPDTLNTIGNFTNNGT 928
 Db 990 KHQTTTHETKGGQGVSTKTSQSDITVAIKEGQPTDNLAMETRAKGS-----QFTSNGD 1044
 QY 929 ABINITQGVVVLGNVINDGDLNITTHAKRNORSIIGDOIINKKGLSNI---TDSNNDAAEI 985
 Db 1045 ISINVE-----NAHYEG-----AOFDAQK--GKTVINAGDGLTLAQATDTHSESQS 1089
 QY 986 QIGGNISQKEGNLTISD-----KINITKQITIKKIDGEDSSSDATSNANLTITKT 1036
 Db 1090 NVNGSANLKVGTTPESKDYGGGFNAGTTHHSKEQTTAKVGTITGSQIELNAGHNLTLQG 1149
 QY 1037 KELKLITEDLSISGFNKAETAKDGRDLTIGNSDNGSCAAEA---KTVTFNNVKDSKISAD 1093
 Db 1150 THLSSEQDIALNATNKKVDLQASSEHTEKGNLSC--GVQAGFGKKMT-----DDASSVN 1202
 QY 1094 GHNVTLNS-----KVTSSSNGGRESNDNT--GLTITAKNVEVKNIDITSL----- 1138
 Db 1203 G-----LGSQAQFAIGKQDEKSVSREGGTINNSGNLTINGSVHLQCAQVNSKDTQLTSQS 1258
 QY 1139 -----KTVNITASEKVTTTAGSTINATNGKASITT-----KT 1170
 Db 1259 DIEITSAQSTDYKNNWGTDIGFNCKKNTNTPKEVTEEPATSIHNGKLLVNVDEQOKT 1318
 QY 1171 GDISCTISGNTVSVSATVDLTITKSGSKIEAKSGEANTV--SATGTIGTIS----- 1219
 Db 1319 SHONATLETGTLTINSNKDLTL-----SG-ANVTADSVTGNVGGSLINIASQKESDR 1368
 QY 1220 -----GNTVNVNTANAGDL-----TVGNAGAEINATEGAATLTATGN 1254
 Db 1369 HVTGVNVGVNHTNPKSSQVNTAKAGGSLLLEKTKIDTIDSGIKSSTDAISDKYNSLSS 1428
 QY 1255 TLTTEAGSSITSTKGOVDLLAQNGSIAGSINAANVTLTGTLTITVAGSDIKATSGTLVI 1314

Db 1429 TIADTKGIS-DETKAKID--QGFGKVGNGIK--NI---VTGAEGHTANADIKVT----- 1474
QY 1315 NAKDAKLNKGDSDSTEVNAVNASGSGSVTAATSSVNITGDNTVNGNLNIIISKGRNTV 1374
Db 1475 -----HVDNDVATKTI-----SLTNNDLNLVNGS-TKLTAIEIVSQGQ--V 1515
QY 1375 RLKGEIEVKYIQPGVASVEEIVAKRVLKVKDLSDEERETLAKLGVA--VRFVEPNPT 1433
Db 1516 DLGGSVSKLENIE-----GHYFAGADLDLSSVVD-----LAKOLVGGDISFSPKVT 1564
QY 1434 -ITVNTQ 1439
Db 1565 NETVNTK 1571
RESULT 9
H83135
probable adhesin PA4082 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83135
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83135
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1018 (Sto)
A:Cross-references: GB:AE004824; GB:AE004091; NID:g9950277; PIDN:AAG07459.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4082
Query Match 7.5%; Score 553.5; DB 2: Length 1018;
Best local similarity 22.6%; Pred. No. 2.5e-15;
Matches 301; Conservative 181; Mismatches 453; Indels 399; Gaps 61;
QY 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGE---KPMKVRHLALPLSALLSL 57
Db 1 MNKYAL-----VNVSGGCVNVVSEGRRRGKPGAKA-----AIASVALL 43
QY 58 GVTSI-POSVLASGLOGMDVHGTATMQV---DGNKTIIRNSVDATII-NWKQFNIDONEM 112
Db 44 GATAPALAPALPS---GGTVVGGSGANGEIHLGGNSLSVNVKQVKDLIANWDSFSAAGER 100
QY 113 VQFLOENNSAVENRVTSTNOISOLKGLDSNGOVFLNPNGITIGKDAIINTNGFTASTL 172
Db 101 VIFNPSSSTALNRVIGTRASDIQGRIDANGOVFLVNPNGVLFGRGAQVNVGSLVASTL 160
QY 173 DIS-----NENKARNFTFEOTKDKALAEIVNH-GLITVGKDSVNLIGKVKNEGIVSN 227
Db 161 DITDAEFNGNSSRYRFTGPTNG-----VLNHGGAITAEAGSIALLGAVDNRGVLAQ 215
QY 228 GGSTSLLAGOKITISDIINTIYTSIAAPNEAV--NLGDIFAKGNNIVRAAT----- 279
Db 216 MGVGLGAGSDLLTLPDGNKLLDIRVDAGVANALASNGGLLKADGGRVLMMAARTANALLN 275
QY 280 --IRNOGKLSADSVSKDSKNIVL- SAKGEAEITGGVISAQ--NQAKGKGLMITGDKVT 334
Db 276 TVVNSQGAIEARSL-RCKNGRIVLDGPDGKVMVGGALSANALNGPCHGCTVEVRGQAVE 334
QY 335 LKTCATVIDLSKGEK--GETVILGDERGEGKNGIQLAKTKTSLEKGSTINVSKEKGGFAIV 392
Db 335 VALGTQVNTLASNLGNTWKIAADKID-----VRPSAVSDGVVVAHDTLSRN----- 381
QY 393 WGDIALIDGNIN-AQSGCDTAKTGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVINAED 451
Db 382 -----LASTNIELVSTKGLDLDGSVWASGNRL----- 410
QY 452 PLFNNTGINDFPTGTGEADSPKKNSLKTTLNTTITSNYLKNATWMTNITASRKLTVNSS 511

Db 411 -----GLGSAADLTNLGRNAS-----CAKAGLELKAEGA 440
QY 512 INIGSNHLLHLSKG-----QRGGVQIDG--DITSKGNLTIVSGGWDVHKRNTITLDQ 564
Db 441 IDI--NDKIVLGGAGSALADAGEHRVNGTASLAGANATYVSGGY-----YTVQ 493
QY 565 FLNITAAVAPEG-----GNKARDAANAIVQAQTVITTEGCKDFRANVSLNCTGKGLN 620
Db 494 LAQLOAINKNLDGLVYVIGN-----ILGGSYCT-----ALQSIGPAG 532
QY 621 IISVNNIHLNLTSGTINISGNITINQTRKNTSYWQTSYQSHSHVNSALNLETGANFTFIK 680
Db 533 VFSG-----TLDGLNSIGNLSIS-----NTGPNV----- 557
QY 681 YISSNSKGLTQYRSSAGVNFNGVNNMFLKEGAKVNFKLKPNENMTSKPLRPLA 740
Db 558 -----GLFA--RSS-----GTLNKLN-----NLRVSDNTYGGSPSLGALV 593
QY 741 NITATGGSVFFDIYANHSGRAELKMSSEINISGANFTLNHSHVRGDDAFKINKDLTINA 800
Db 594 GT-----NSGRIANVYASGVSV--VGSRLRSNALGG---LVGRNISQOI 632
QY 801 TMSNFSLRQTKDDFYDGYARNAINSTYISILGGNVTLGGQNSSSSITGNITIEKA-ANV 859
Db 633 ANASVSGGV-----GYAAS-----TAVGG--LVGENFTTAMGPEAVIENAHNV 675
QY 860 TLEANNAPNOQNI RDRVTKLGLSLLVNGSLTGENADIKGNLTISESATFKGT---RDT 916
Db 676 HVAAGSTERN-----LGGVGLVGLNA--KGMIRASGS---QGVETYPG 717
QY 917 LNIITG--NFTNNGTAEINITQGVYKLVNVDGDLNITTHAKRNORSIIIGDILNKKSL 974
Db 718 LNVGLGVNMFHVSVDASAGQVEAGAGNTGVLGLSSGGEIPRQASQSVYKSG-- 775
QY 975 NITDSNDAEIOIGNISQKEGNLTISDDKI--NITKOITIKKIDGIDGSSDSDATSNANL 1032
Db 776 -----LATGLIKRAEGNMLGNLAKASGVTDQ-----GGADLGLGVNNSQS 818
QY 1033 TIKTELKLTEDLSISGNPKAEITAKGDRDLTIGNSDGNSGAEAKTVTFNNVNDKSLA 1092
Db 819 ALETAE---ATPKVSGGNSRVGGLIGHNL-----GGSAHAISRQDV----- 859
QY 1093 DGHNTVTLNSKYKVTSSNGGRESNDSNDTGLTITAKNVEVNDITSLKTVNITASEKVT 1152
Db 860 GGFNSLVGLV---GHNGG-----ELVNDASGRVSAA 889
QY 1153 AGSTINATNGKASITTKTGDISTISGNTVSVSATVDLTTSKSGKIEAKSEANVTSATG 1212
Db 890 ASASV-----GGLVSNAGSILSARSSSTVNGSGRSRIGGLVGENQI---QG 933
QY 1213 TIGGTTISGNTVNVATANAGDLTVNGAEINATEGAATLTATGNTLTTEAGS-SITSTKGV 1271
Db 934 RIVSSMSECTV---SGDYVYVSMG-----GLAGL-----NLGSIYSGVSGKI 972
QY 1272 DLLAQN--GSIAGS 1283
Db 973 DFKPOSHYQIYGA 986
RESULT 10
B99789
hemagglutinin/hemolysin-related protein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B99789
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B99789
A:Status: preliminary

Query Match	7.2%;	Score 531.5;	DB 2;	Length 1268;
Best Local Similarity	22.3%;	Pred. No. 2.7e-14;		
Matches 351;	Conservative 209;	Mismatches 513;	Indels 499;	Gaps 78;

Db 708 QNSSGSGTMDNNATGFIASDK--NLSLEVYNS-----LNTYGMTSGKGDVDVTYNN 756

QY 895 ADIKGNLTIS-----ESATFKG--KTRD-----TLNITGNFTNNGTAE-----INI 933

Db 757 GNLNRRNTIAAEKGDJDAALANGTENMKDITASAGGDLTMTNRRHTVNSNSNVGQNTIVNA 816

QY 934 TQGVYKLGNTVTDGDLNITTHAK--RNRQSIIG-GDIENKKGSLNITDSSNDABIQIGNI 991

Db 817 VNDNNRGNIVDADLVNVTTKGKLYNLYVMYGVYGIALSANSV-----ANNATTIATGD 872

QY 992 SQKEGNLTISDKNITKTIKKGTIDGEDSSDTSNANLTIKTKELKLTEDLSTSGFN 1051

Db 873 -----DVTL-----LIDSKGVGNRGNLHALGVLSKGNLNL--NDNGEIRGY 912

QY 1052 KAEITAKDGRDLTIGNSDNCNSGAEAKVTENNWKSKISADGHNVTLNSKYKTSSSNGG 1111

Db 913 -----DVTL-----ALTCNVDSYKGSGLTSETG--DVTLTANI----- 942

QY 1112 RESNSDNDTGLITAKNVEVN-KDITSLKTVNTITASEKVTITTAGSTINATNGKASITTK- 1169

Db 943 ----VDNAYGL- IAGENVSDAKSTIYNTALTAANKKLIVNAGNLENRDNGNFERNNG 997

QY 1170 ----TGDLSGTISGNTVSVSATVDLTITKSGSKEIAKSGEANYVTSATCTIGT-----LSG 1220

Db 998 ALFGITDNGVGIGKEGVTLISA--QNVYNNSSTIIAENPLNLLS--RGLTDNTRALLSSG 1054

QY 1221 ----NTVNTANAGDLTVNGAEINATEG-----AATITATGNTLTITEAGS 1262

Db 1055 ADALIRAAGTFYNNYATYTSAGNLDVYAASLNASDGRLEDNTATGVIA SDKNLDLSVDN 1114

QY 1263 SIT-----STFKGVDLLAONGSI-----AGSINAANVTLNTTGTLITTVAGSDIKA 1307

Db 1115 SVTVNGISGKGDVHFNVKGLYNRNIAADNALTINALNGVENFK--DIVAGTALT 1171

QY 1308 TSGTLVINA KAKLNGDASDSTEVNAVN--AGSGSVTAATSSVNTIGDLNTVYNGLNI 1366

Db 1172 DTQKYVTN----NNSNMLGQTTAINAVDNNRGNIVGDVSLGVKVTG--NIYALNML 1225

QY 1367 SKQGRVTPLRGKEIVKYIQFGVASVEEVIEAKRVLEKVKVDLSDEBRETTLAKLGVS AVR 1426

Db 1226 SY-----GVAGV-----SANKVTNSGKD-----AVLGGFVGL 1252

QY 1427 FVEPNNTITVNT 1438

Db 1253 ALGANETDWTGT 1264

RESULT 11

E85649

hypothetical protein Z1542 [imported] - Escherichia coli (strain O157:H7, su

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: E85649

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, K.

iller, L.; Grotbeck, E.J.; Davis, N.W.; Llin, A.; Dimalanta, E.; Potamousis,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85649

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1270 <STO>

A:Cross-references: GB:AE005174; NID:g12514410; PIDN:AAG55657.1; GSPDB:GN001

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

C:Gene: Z1542

```
Query Match          7.2%   Score 531.5; DB 2; Length 1270;  
Best Local Similarity 22.3%; Pred. No. 2.7e-14;  
Matches 351; Conservative 209; Mismatches 513; Indels 499; Gaps    78;  
  
QY      46 ALKPLSAMLGLSGVTSIPVSLSAGL--QGMDVHGTATMQVGDKTKIRNSVDALIN-W 102  
|||:::>:::  
||||:
```

Db 15 ALTPIALMM-----LSFPVASOAGLVIRKNTVYANGVPVVDINKP-----NGSGLSHNIW 67
QY 103 KQBNIDONEMV-----QFLOENN-----SAVENRYTSNISOLKGLDLSN 143
Db 68 DNLNVKGVVFNNSANESTSLAGNIQSNLTSSAKVILNEVTSKPNSTINGMEVA 127
QY 144 G---QVFLNPNGITTKDAINTN---GFTASTLDSNENIKARNETPFTQTKDKALAEIV 198
Db 128 GKRADLIANPNGITVNGGSGINTGKLTLTTPDQD-----DKLAGYSV 173
QY 199 NHGLITVKGDSVN-----LIGGKVKNEGVISVNGSGISLLAGQKITISDIINPTI 249
Db 174 NGTITLGLKLDNASPTIELSRNVVAVKVSADENLVAGNYYNAAGQ-----V 222
QY 250 TYSIAAPEN-----EAVNLGDIIFAKGNI---NVRAATIRNOGKLS---ADSVSKDKSGNI 299
Db 223 TGSVATGSRNGYSVDVAKLGGMYANKISLVSTKGVGRNLGVIAGGVNGYSIDSKGNL 282
QY 300 V-----LSAKEGEAEIGGVISAQNOQ-----AKGKLMITGPKVTL 335
Db 283 LNSNAQIOSASTINLTNGTLDNTTGTVTSGTISLNTNKTIVNTRAGNISTMGD-IVV 341
QY 336 KTGAVIDLKSGEGEYVLGD-----ERGEKN-GIOLAKTSLEKSTINVSKEKG 387
Db 342 NSGTIDNTNGKLAAGMLAVDTNNTATLNSGKSSVGIE-AGLVALKTGLTNNNGQIRG 400
QY 388 GFALVWGDIADIGNTNAOGSDIATGGFVETSGHDLFTKDNATYDAKEWLLDFDNVSI 447
Db 401 GY-----VGLSAALN-NNNGDLOITG-----DIALISNGVD-----432
QY 448 NAEDPLENTGINDPEPTGCE-----ASDPKNSLKTTLTNTTISNYLKNWMTMITAS 503
Db 433 -----NNKGL---IRKSTGHVIGAAGSVNNGSTKTDATGSSDS-----LGIIA- 473
QY 504 RKLTVSINIGNSHLILHSKGORGGOVDIGDITSKGNLTIYSGGWVHVHKNITLDO 563
Db 474 -----DTGVIGAN-----NINNGQ---IASNGVSLSSYSTID- 506
QY 564 GFLNITAASVAFEGGNKARDAANAKIAQGTVTITGEKDFRANVNSLGTGKGLNITS 623
Db 507 -----DYAGKILSKSVIKG-----S 523
QY 624 SVNNLTHNLG-----TNISGNITIN-----QTRKNTSVMTQSHDGHVNSALNLETGAN 675
Db 524 SLRNDTGGISGKOGIEVAVGSLTNIGVISSEEGDISLLANSVDNHG-----GFMMGQN 578
QY 676 FTFKYISSNSKGLTTQYRSAGVNFNGVNGNMSFNLKEGAKVNFKLKPNENMTSKPLP 735
Db 579 ITM-----ESMSGVNNNTALIVASKKLKINARGSIENROCNFGNAYGL- 622
QY 736 IREFLANITATGGSVPFDIYANHSRGAE-LKMSIENISNGANFTLNSHVGRDDAFK 794
Db 623 ---YFGPQQTGGWV-----GKEIUSGONIYN-----NSRLIAEDG-----658
QY 795 DLTNATNSFSLR-----QTKDDFYCYARNAINSTYNTSIILGNTVLGG---840
Db 659 PLTLOAONTEDNTRALVTSADASIQVGGTYNNYA-----TTWS-----AGNLDIDATTL 709
QY 841 QNSSS-----STGTNTTIEKAANTVLEANNAPNOQNIRDRVIKLGSLLVNGSLSTGEN 894
Db 710 QNSSSGTMDNNATGFTASDK--NLSLEVNS-----LNYMGWISGKGDVDTVNN 758
QY 895 ADIRKGLTIS-----ESATFKG--KTRD-----TLNITGNFTNNGTAP-----INI 933
Db 759 GNLVNRNTIAAEKGLDIAALNGIENWKDIASAGDLTMTNTRHVTNNSNSNMGQNVINA 818
QY 934 TOGVVKGNTVNDGDLNITTHAK--RNORSITIG-GDIINKKGSNITDSNNDABIQGGNI 991
Db 819 VNDINNRGNIVSDADLVNTTKGNLYNLYVMVGDIALSANSV-----ANNNATIEATGDL 874
QY 992 SQKEGNTLISDDKINIKQTIKIGIDGSSDATSNANLTIKTKELKLTEDLSLGEN 1051
Db 875 -----IDSKGVNNGRNLHALNGVLSVKGNNLN-NDNGEIRGYG 914

QY 1052 KAEITAKDGRDLTIGNSNDGNGSAGFAKTVTFNNVKDSKISADGHNVTLSNVKVTSSSNGG 1111
Db 915 -----DVL-----ALTNVDYSYKSLTSETG-DVILTANI-----944
QY 1112 RESNSDNTGLTITAKNVEVN-KDITSLKTVNITASEKVTITAGSTINATNGKASITTK- 1169
Db 945 -----VDNAYGL-IAGENVSVDASTIYNNTALIAANKKLVINAGNLENROGNFLNRNG 999
QY 1170 -----TGDISGTISNTVSVSATVDLTTKSGSKIEAKSGEANVTSATGTIGT-----ISG 1220
Db 1000 ALFGITDNGGVIVGREGVTLSA--QNVYNNNSIIAENGPLNLLS-RTLONTALRLSSG 1056
QY 1221 -----NTVNTANAGDLTVNGAEINATEG-----AATLTATGNTLTUTEAGS 1262
Db 1057 ADALIRAAAGTFYNNYATTYSAGNLDVYAASLNNASDGRLEDNTATGVIASOKNLDLSVDN 1116
QY 1263 SIT-----STGQVLLAONGSI-----AGSINAANVTLTNTTGLTTVAGSDIKA 1307
Db 1117 SVTYNGWISGKGDVHFNVKLTLYNRNIAADNALTINALNGVENFK---DIVAGTALTI 1173
QY 1308 TSGTLVINAKDAKLGNDASGDSTEVENAVN-ASGSGSVTAATSSSVNITGDLTNTVNLNII 1366
Db 1174 DTQKYVTN-----NSNMLGQTIALNAVNDINNRGNIVGDYSLGVKTTG--NIYNYLNL 1227
QY 1367 SKDGRNTVLRGKEIEVKYIOPGVASVEVEIAKRVLEKVDLSDEERETLAKLGVSAVR 1426
Db 1228 SY-----GVAGV---SANKVTNSGKD-----AVLGGFYGL 1254
QY 1427 FVEPNNTITVNT 1438
Db 1255 ALEANEETDGT 1266

RESULT 12
T31102
filamentous hemagglutinin 1 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31102
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326; PMID:9811662
A:Accession: T31102
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4152 <WAR>
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1
C:Genetics:
A:Gene: lspA1

Query Match 7.1%; Score 523.5; DB 2; Length 4152;
Best Local Similarity 20.1%; Pred. No. 2.6e-13;
Matches 399; Conservative 239; Mismatches 636; Indels 711; Gaps 83;
QY 2 NKIYRLKFSKRNALVAVSELARGCDHST-----EKSEKPARMKVR-----44
Db 3 NKRYKLIFSFKVKNCLVPVAENTKASGNSGSSNSKIAEDQDEEEDPSLACSLSPLSSSTH 62
QY 45 LAL-----RPLSALLSLGVTSPQSVLASGLQGMVDVHGVTATMQVDGNKTI-- 92
Db 63 LGLHNSPLKVFEGKSLSVLLSL---MPATPLLAQQNYAEALNGKVVVDSDSHSSTRIY 118
QY 93 -----RNSVDATII-----NWQFNIDQENEMVQFLOENNNNSAVFNRTVSNQ 132
Db 119 EQKTDNDSKDGIVVVEIANPEVDGVSDNRFKEFNIP-----NSAVFNNSRTES 166
QY 133 ISOLKGLDLSNGQ-----VELINPGITIGKDA 160
Db 167 TSOLVGKLIHANTLOKEAKLIILNQVTGDHESNIOGALEAVAGKKADLIIVNPNGLTNGVK 226
QY 161 IINTNGFTASTLDI--SNENIKARNFTPEQTKDKALAEIVNHGLITVKGDSVNLGKGV 218

Db 227 TINTDFVVSSTDIIPHREN-----GLLSV-RNGKVTI----- 258
QY 219 KNEGIVSVNGSISLACQKTIISDIINPTTYSIAAPENE-AVNLGDI-FAKGG--NI 273
Db 259 -DKGVAATNGLSHFVFAVARNIDOKGI-----TVAKTENOKSVNPANITFAAGSLNLYN 311
QY 274 NVRAATIRNOG-----KLSADSVDKSGCNI--VLSAKEGEAEIGGVISAQNOQAKG 323
Db 312 KTREATPISSTGSRSTSDPAISADSAGSMYGSNIKFFVVDKAGGVKHKGIIFSENDI-- 368
QY 324 KGLMITGDKVTLK--TCGAVIDLSGKEGETYLGDEREGKNGIOIAKKTSLPEKGTIN 380
Db 369 -NIKMDGNGASLKELYAKKDILAKD-----IELTEKQLOANNKII 410
QY 381 VSGKEKGFAIWGMDIALIDGNIINAQSGDIAKTGGFVETSGHDLFKDAIYDAKEMLL 440
Db 411 LMS-----TKINLRNASEVS-----ADNVNKSLENAL 439
QY 441 DFDNYSINAEDPLFNNTGIDNFTGTGEASDPKKNSEL-----KTL--TNTTIS 489
Db 440 ENASASANSLOVIVTKIEVNRSSKVSAGTANIKASNTITLGGSSVANKITLNTNNATLN 499
QY 490 NYLK-NAWTMITASRKLTVNASSINIGSNHILHSGQGGVQIDGDTSGKGNLTII 548
Db 500 NOSKLSAKDMELNTHNITLNT-----SKLSAQKANIKTENTLNGEASLVAEKLDIN 553
QY 549 SGGVVD-VHKNIITDQGLNITAAVAFEGGNKARDAANAIVAQGTIVITGEGKDFRA 607
Db 554 A--IDKTNNGTITAGLTANITTKAL-----ENRD--NALILAHQNLNFTVNGSHY-V 600
QY 608 NVSLUNGTKGLNITISSVNNTHLNGSITNITNTRKNTSYWQTS--DSHWNVS 666
Db 601 NKGDIVSKDAIVTFSNNSDPTSGSLVDAQNLLTVN--NNFNITOGSEIILHGNVT 657
QY 667 ALNLETGANFTFIKYSISNSKGLTQYRSSAGVNFNGVNG-----NMSFNLKEG 715
Db 658 ---LNAKGNFT-----NSGNLTMKR-----LNISIESFINAGNLTGKLEVHSNTT 703
QY 716 AKVNFKLKPNENMNTS-----KPLPIRELANITATGGSV-----FEDIYAN 757
Db 704 VKNQGLKLSIENLISSTKDTFTNGTLLGLEALKIASGGNFTNASNGSLASNSLSDIYN 763
QY 758 H-SGRGAELKMEINISGANFTLNSHVRGDDAFKINKDLTINATNSFSLROTKDDFYD 816
Db 764 NFTNNGTIESVKSLNITNNTYFINNATIKSYGVNLI-----TSQGNF-----TNSNGT 812
QY 817 GYARNAINSTYNIISLGNVTGLGONSSSITGNIT----- 852
Db 813 VMSHDLNITSOANILNKLLAGQGLNLTAQGNITNDSNSTALAVLHSNNDINLANNK 872
QY 853 -----IEKAANYTLEANNAPNOQNRDRV--IKLGSLLV-----G 886
Db 873 VYNGEYISQAGNISVEAKLLHNDVKLGNITTTKSGNATVKTNSIGGGLHDANSIRVG 932
QY 887 SLSUTGENAD-----IKGNLTFISSEATFKG-----TRDTLNTIG-- 921
Db 933 ELTLNGKPADLDNOLKVALRGIYAGSNLTFKAKEGEKEQKSTAQAKIINRGITNVKNKL 992
QY 922 NFTNNGTAETNITOGVVKLGNVTNGDNLNIT-----THAK-----RNQRSITIGDILNK 970
Db 993 EYGSNVVDVNNMRSMQVNLVEKIFNGDNPIITLTKNGCVTFKDFSNRRRRASNDGECTNK 1052
QY 971 KGSNLNIT-----DSNNDABEIQGGNISQKEGNLTITSSDKINITKTIKKGID-- 1018
Db 1053 KTFDNVAHLIEEAFSGYSNGNDHRASDDGHVKSPPYLLVLAQAVNTEGENYKLTALQHI 1112
QY 1019 -GEDSSSDATSNANLTIKTK--ELKLT-EDLSISGFNKAEIT-----AKDGRD---LTTG 1066
Db 1113 FGPWNLDLTINNDTTLNDKWNQULKKEFKKNNGENHNSININIVPADEGVKAKIFAG 1172
QY 1067 NSNDGNSGAETVTFNNVKDSKISADGHNVTLNSKVKTSSSNG-----GREN 1115

Db 1173 VLNRNGTNGVEDKVIYQELNDKAKKEYEDKFAKKPQGRFQNGFQWAGDWAKEGNESY 1232
QY 1116 SONDITGLTITAKNVEVNDKIDTSLKTNITASE-KVTTTAGSTIN----- 1158
Db 1233 GSKET-----EEKYNGIKKEHTVNIQKHEIKVPTVSFENLNINHQDDKSDGDKSI 1284
QY 1159 -----ATNGKA-----SIIT 1168
Db 1285 ISELLAQPIYVAKADVPDPRVAQNDKAVDEGLYRTRLSYINQNNYLGAKYFFNQLDT 1344
QY 1169 KTGDISG-----TISGNTVSV-----SATV---DLT 1191
Db 1345 EDDKLGKIRIGDNYFEHQLITRLEKVDADNHLTLKHLHDIALVKLLDSASIOAKDLN 1404
QY 1192 TKSGS-----KIEAKSGEANVTS-----ATGTI-- 1214
Db 1405 LKVGELATKEQDNLDKEDIVWYVYKTEVNAQEVLLVPOVYLAKOTIEEVKQRGVGTQIRA 1464
QY 1215 -----GGTISGNTVNVTA-----NAGDLTVNGAEINATEGAATLTATGNTLT 1257
Db 1465 GIIDVKVDDVRNVTGTITAGYAVGLEAKNKLNKTGDIL---SRLSKLVGKKGLESTGVTVV 1521
QY 1258 TEAG-----SSITSTKGQV----- 1271
Db 1522 DETGATKVRKARIKSEGHYILETDKDNVDLTASELKGNTGQIKAKDLNLDIYETSYK 1581
QY 1272 ---DLIAQNG-----SIAGSIN--AANVTLNTT----- 1294
Db 1582 KYEKLFGKNGEIGDRVQTTSQAKSVGTDSFHLHLSLEGDVNTGSLKANRTTGVVK 1641
QY 1295 GTLTTVAGSDI-----KATSGTLVINAKDAKLNGDSGDSTEVAVNASGSGSVTAATS 1348
Db 1642 GDFNTKAGKDLFHRQIDTIVTSQT--VYSASASGGQSAGISLTDQGVETYNKKTATAGAN 1699
QY 1349 SSV-----NITGDLNTVNG-LNIIISKDGRNTVRLRGKEVEKYIOPGVAS 1392
Db 1700 ADVTNFMKRTRETETSLTHRNSEFNALSGLYVMGKADIGGVDI-NRDVEV-----IKT 1752
QY 1393 VERVIEAKRVLKVKDL---SDEERETLAKLGVSAVRFVPEPNTITVTFQNETTRPSQ 1449
Db 1753 PEERIAEORAAEAKAEVKEANEASATAK-----ETEEAENDVNAEKDKIRPKFK 1803
QY 1450 VIISE 1454
Db 1804 KLJDE 1808

RESULT 13
B81192
hemagglutinin/hemolysin-related protein NMB0497 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81192
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, B.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: B81192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1975 <TET>
A:Cross-references: GB:A8002406; GB:A8002406; NID:g7225720; PIDN:AAF40929.1; PID:g722 A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0497

Query Match 7.0%; Score 520; DB 2; Length 1975;
Best Local Similarity 22.7%; Pred. No. 1.4e-13;
Matches 384; Conservative 227; Mismatches 551; Indels 528; Gaps 88;

QY	1	MNK - IYRKFSCRRLNALVAVSELARGCDBHSTKGESEKPARMKVR-----HLAKPKPLSA	52
Db	1	MNKGHLRIIFSKKHSTMWAAETAANSOGKGGKOAGSSVSUKTSDGLCKLTKTLTKLVLC	60
QY	53	MLLSLGVT-----SIP---OSVLASLGQMDVVHGTAITMQVDGNKTIIIRNSVDA	98
Db	61	SIVLSJMWLPAPHAHQTTDKSAPKNQOVVILTKNTCAPLVN-----IQTNGRGLSHN----	112
QY	99	IINWKOFNDQBHEVMQFLOENNNSAVFN---RVTSNQI-----SQLKGILDNSGO---VFL	148
Db	113	--RYTOFDVNDKGAVLNDRNRNPFWVGSAOLINEVGRTASKLINGLTVTGOKADVII	170
QY	149	INPNGITIGKDAIINTNGFTASTLDITSNIENIKARNFTEQTKDKALAEIVAHGLIT----	204
Db	171	ANPNGIT-----VNGGFG-----ITGSGRILUTTGAP	196
QY	205	-VKDGSVNLIQKGVKNEGISV-----NGG-----SISL---LAGOKITIS--	242
Db	197	QIQKDGA--LTGFQDVR-QGTFLVGAAGWNKDGADYTGVLARAVALOQKLOGKNLAVSTG	255
QY	243	----DIINFPTINYSTAAPENEAVNLGDIFAKGN-----INVRAATIRNOCKLSADSYSK	293
Db	254	PQKVQDVASGEISAGTAAGTKPTIAL-DPAALCGDSITLIANEKGVGVKNAGTLEA-----	307
QY	294	DKSGNIVLSAKBEAGEAIGGVISAONOAKGGKLMTIGDKVTLKTKGAVIDLSKEGETYL	353
Db	308	-----AKO-----LIVISSGRIENGSHIATTAD-----GTEASPYYL	339
QY	354	GDEREGKNGIOLAKKTSLEKSTINVSGBKPGGFAIVWG-DIALIDGINAQSGDIA	412
Db	340	SIEETTEKGAAG-----TFISINGRI-----ESKGLLVITGEDISLRNGAV-VQNNG-S	386
QY	413	KTGGFVETSGHDLFIKDNAIVDAKEMWLDFDNVSVNAEDPLF-----NNTGIND-BEPTGT	467
Db	387	RPATTVLNAGHNVLIESKTNV-----NAKPATLSADGRFVIKEASIQGT	433
QY	468	GEASDPKKNSEL-----KTLTNTTISI-N-YLKNAWTMNITASRKL-----TVNS	510
Db	434	TVSYSSKNGAELGNTRITGADVTVLSUNGTISSSAVIDAKDTAHEAGKPSLEASTVTS	493
QY	511	SINIGSNHLILHSKQRGGVOIGDITSK-----GGNLITYSGGWVDVHKNTILDOQ	564
Db	494	DIBLNGS-----IKGKQALLADDNIATAKTNTLNTPGNLVHTG-KDLNLNVYDKD-	544
QY	565	FLMITAASVAFEGGNKARKANAAKIVAQGTVTITGEGKDFRANNVSLNGCTGKGLNISS	624
Db	545	--LSAASIHLKSDN-----AAH-----ITGTSKTLTASK-DMGVEAGSELNVTNT	585
QY	625	VNMLTNLSCGTINI-SGNITINOTRKRTSYWQTSDDSHNVSAALNLETGANETFFIKY	681
Db	586	--NLRTN-SGNLHIQAAGNIQL-RNTKLN-----AAKALETTA-----	620
QY	682	ISSN--SKGLUTTQYSAGVNFNG-----VNGMSFNLEKGAQVFKLPNPMNMNTSKPL	734
Db	621	LQGNIVYSDGL-----HAVSADGHVSLLANGNADET-----GHNTLTAKA-	659
QY	735	PIREFLANITATGCVSFDFIYANHSGRGAEKMKSEINISGANFTLNSHVHRGDDAFKINK	794
Db	660	-----DVNAGSVGKG-RLKADNTNITSSG-----	683
QY	795	DLFINATNSFSLROTKDDFYDGYARNAINSTY-NISILGGNVTLGCQN-SSSITGNIT	852
Db	684	DITLVAG-----IQUGCKQRNSINGKRHISINNNGGNADLKNLVHAKSAGALNTH	735
QY	853	IEKA---ANYVTLEA--NNAPNQONIR-----DRVKILGSLLL	883
Db	736	SDRALSENTKLESTINTHUNAQHREVTLNQVDAYAHRHLISITGSOIWONDKLPSANKLV	795
QY	884	VNGSLSTGTENADIKGNLTITISESATFKGTRDTLNTIGNFTNNGTAEBINITQG--VVKUG	941
Db	796	ANGVALNLARYSQIADNTTLURAGA-----INLTAGTALVKRG	832
QY	942	NVTNDGLNLTTHAKRNORSIIIGDIIINKKGS LNIT-----DSNNDAEIQIGGN--IS	992

Db	833	NI--NNSTVSTKTLFEDNAELKPLAGRLNITPAGSGVLTTPANRISAHVNDLSIKTGGKLLLS	891
QY	993	QREGNLTITSSDKINTKQITTKKGIDGDS--SSDATSNANMLTKTKVELKLTDELDSISGF	1050
Db	892	AKGNAGAPSAQVSLEAKNGNIRLTGTDLRGSKITAGKNLVATTGKGLNIEAVNNSE	951
QY	1051	NK-----AETAKDGRDLTIGNSDNGNSGAEAETFFNVKDSKISAGCHNVVTLNSKY	1103
Db	952	SNYFPTQKAAELNQR-SKELEQOIAQLKSSPKSLIP--TLOBERDLAFYIOAINKEY	1008
QY	1104	KTSSNG-----GRESNSDNDTGLTITAKRNEVN-KDITSLKTVNITASEKVTITAGSTI	1157
Db	1009	KGKKPKGEYLOAKLSAQNID---LISAQGEISGSDITASKKLNHAAAGVLPKADSEA	1065
QY	1158	-----NATNGKASITTKTG-----DISGTISG---	1179
Db	1066	AAILDIGITDQYEIGKPTYKSHYDKAALNKPRLTGRGTGVSIAHAAALDDARIILIGASEI	1125
QY	1180	-----NTVSSANVDTLTKGS-----KIEAKSG-----EANY	1207
Db	1126	KAPSGSIDIKAHSDIVLEAGONDAYTFELTKGSGKLIIRKTKFTSRDHLIMPAPVELTA	1185
QY	1208	TSATGTIGTISGTVNTVYANAGDITVNGAEN--ATEG-----AALTATGNTLTTTEAG	1261
Db	1186	NGITLQAGNGNIENTRFNAPAGKVLTVAGBELQALLAEGTIKHIELDVOKSRFIGIKVG	1245
QY	1262	SS-----TTSYKQGVDLLAONGSTAGSINANVTINTTGLTTVAGSDIKATSGTLVIN	1315
Db	1246	KSNYSKNELNETKLPVRVVAQT---AATKSGMDTVLEGTGKTLIAGADIQAGVEKA---	1300
QY	1316	AKDAKL-----NGDASGDSTEVAIV--NASGSGV--TAATSSSVNITGDLMTVNGLN	1364
Db	1301	RADAKILKGIYVNRIOSEKLETNSTVWOKQAGRGSTIETLKLFSFESPTPKLTAPGGY	1360
QY	1365	I--ISKDGRNTVLRGKEIEVYIOPGVASBEVTEAKRV----LEKYKDLSDERETLA	1418
Db	1361	IVDIPKGNLKT-----EIEKLAKOPEAYLKQLQVAKNVNMNQVLAIDKYDKWYKQEGLT	1414
QY	1419	KLGVSAVRFV	1428
Db	1415	RAGAAVTII	1424
RESULT 14			
AC0304			
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain C092)			
C:Species: Yersinia pestis			
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001			
C:Accession: AC0304			
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; P.			
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.			
II, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead,			
Nature 413, 523-527, 2001			
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.			
A:Reference number: AB00001; M0ID:21470413; PMID:11586360			
A:Accession: AC0304			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-2535 <KUR>			
A:Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00			
C:Gene: YPO2490			

	Query Match	7.04;	Score 520;	DB 2;	Length 2535;	
	Best Local Similarity	19.7%;	Pred. No. 1.9e-13;			
	Matches 385;	Conservative 272;	Mismatches 656;	Indels 642;	Gaps	85;
Qy	1 MNK -IYRLKFSCRNALVAVSELARGCDHISTEGKSEKPAKMVRHLAKPLGS -AMLSLIG 58					
	: :	:	:	:	:	
Dd	5 MNKNLYRIIFNKRGMVMIVADIAASGRASSPSSGLGHTOHRIRTSALSTSFSLLLALG 64					
Qy	59 VTS -----IPQSVLASLGQQMDVV---HGT-----ATMQVDGNKTI 91					

Db	1096	LNNRGSDYVLQQLNNDPVIKFR-----LCDNAYEORLVRDQVLALTQCAVADSY--RSA	1149
QY	811	KDDFYDGYARN-AINSTYNTISILGNGVNTLGGQNSSSSITGNITIEKANVLTLEANNAPNQ	869
Db	1150	QEQEALFAAGLEYSKAFNIA-LGTHLS-----AEQMAALTHNI-----VLMETRDVAGQ	1198
QY	870	QNIQDRV-----IKGLSLLVNGSLSLTGENADI-----KGNLTISSEATFKCKTROTL	917
Db	1199	TVLVPVVVLAVKVPQDQLQANGAL-IAAENISLTEVQGFNAGATATNDLKISMAQDITL	1257
QY	918	NITGNFTNGTAENITQGVVKLGKLVNTDGLNITHAKRNQRSI---IGGDIINKKGS	974
Db	1258	NNRGELLQAG-----GDMQLSTLNSDIDL---TSARINATNLQDSGRDVLIRTDISA	1306
QY	975	NITDSNDAEIQIGNISQKE-----GNLTISSEKINITHQITIKKID-----GED	1021
Db	1307	QLSSDN-----GAVSRDQITILGLA-----SINVSNNAATINGRFOFMOGASLVNGOD	1354
QY	1022	SSSDATSENAL-TIKYKEKLIEDISIGFNKAEITAKD---GRDUTIGNSDNGSGAE	1076
Db	1355	LQVTTGGWQLETVQTRDQLSHD-----GRGSATSESHIRHLGSEVNVGALTAN--VD	1406
QY	1077	AKTVTFNVVKDKSISADGHNVTLNSKVK-----TSSNGGRESNDNDTGUTITAKNVEVN	1132
Db	1407	NLTAVGANINAAFLVYQAOINISLSAATDSLHVTESSSKRHTSSVNLVDETLGCSQLNAT	1466
QY	1133	KDITSLKVNITASEKVTITAGSTINATNGKASITTKT-----GDISGTIS	1178
Db	1467	GDINLQAQDITILRASAVQTDGALLAAGDGVLLTQTQEHQDEORNHITGLSKIASSTLT	1526
QY	1179	GNTVSVSATVDLITKSGSIEAKSEANVTSAITGIGT---ISGNTVNVTVANAGDULTVGN	1236
Db	1527	RTEDSLSQTLAV---GSMLSA--GSDIVSGKNIAVMGNSVVAODDLSLRAQE-NITVGT	1579
QY	1237	GAE-----INATECAATLTATGNTLTTEAGSISINSTRQOVDLLAQNCSIAQ	1282
Db	1580	AQOSESESHLFQKKSGLMSTGGIVTCVGSSTKMTDSGQSISV-----GSTVC	1629
QY	1283	SINAAVNTLTGTLTITVAGSDIKATSGTILINAKDAKNGD-----ASGDSTFVNAYNV	1336
Db	1630	SV-----LGNYSMTAGEDLR-VQGAEEVLAQRDLNLTGKNVSIILAEENQUTQSHTV	1679
QY	1337	ASCGG-----SVTAATSSSVNITGDLNVTNGLNLIISKDGNHVRVLRGKEI	1381
Db	1680	OKOSGLTILASGAVGSVNAVNTAVTAKAASESSGRGLAQG-----	1720
QY	1382	EVKIYIQPGVASVE---EVIEAK-----RVLEKVKOLSDSEERETLAKLGVSAREV	1428
Db	1721	-KAAALNGVQAVQAGLVQAGGGDAASMFGISAGLSQKSSSQHGEQHTVGTST----L	1775
QY	1429	EPNNTIVTQNEFTTRPSSQVITIESEKACFFSSGN	1463
Db	1776	TAGNNITINATGEGNAANSQDIV--QGSQLOAGGD	1809

G81044 hemagglutinin/hemolysin-related protein NM1779 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81044
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Neilson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigiani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58:
A:Reference number: AB1000; UID:20175755; PMID:10710307
A:Accession: G81044
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 11-1995 <TET>
A:Cross-reference: CB16000027, CB16000000, CB16000000, CB16000000

A: Experimental source: serogroup B, strain MC58

C: Genetics:
A: Gene: NMB1779

Query Match		7.0%	Score 517.5	DB 2	Length 1995
Best Local Similarity		22.8%	Pred No. 1.8e-13		
Matches		385	Conservative 228	Mismatches 555	Indels 519
					Gaps 87
QY	1	MNK-IYRLKFSKRLNALVAVSELARGCDHSTKESKSEKPMKVR			HIALKPLSA 52
DB	1	MNKGHLRIIFSCKHSTWVAETANSOGKQAGSSVSLSKTSGLCGKLTTLKTLVLC			60
QY	53	MLLSLGVY-----SIP-----QSVLASGLOGMDVHGTATPMQVDGNKTIIRNSVDA			98
DB	61	SLVLSMVLPAHAQITTDKSAPKNOQVVIKNTGTGAPLVN-----IOTPNRGRLSHN-----			112
QY	99	IINKQFNIDONEMVQFLOENNSAVFN-----RVTSNOI-----SOLKGLDSNGQ-----VFL			148
DB	113	--RYTQFDVNDKNGAVLNNDRNNPFVVKGAQLILNEVRGTASKLNGIVTVGGOKADVII			170
QY	149	INPNGIITIGDAIINTNGFTASTLDISNENIKARNFTFEQTKDKALAEIVNHGLIT-----			204
DB	171	ANPNCIT-----VNGGCF-----			196
QY	205	-VGKDGSVNLIGGVKKNVGVISV-----NGG-----SISL-----LAGOKITIS--			242
DB	197	QIGKDGAA--LTFEDVR--OGTLTVGAAGNDKGGADYTGVLARAVALOGKLOGKLAIVSTG			253
QY	243	-----DIINPTITYSIAAPENAEVNLGDIKAGGNNINRAATIRNOGLSADSVS--KDKS			296
DB	254	POKVDYASGESISAGTACTGPTIAL-DTAAALGG-----MYADSTILIANEK			298
QY	297	GNIVLSAKEGEAEIGGIVISAQNOQAKGKLMITGDKVTLKTGAVIDLSGEGGETYLGGD			356
DB	299	GVGVNAGTLEAAKQILVTSGRISNGRIATTAD-----GTEASPTLYSIE			345
QY	357	ERGGKNGIQIAKTSLEKSTINVSQKKEGGFAIVNG--DIALIDGNINAQSGDIAKTG			415
DB	346	TEKGAAG-----TFISNGGRI-----ESKGLLVETGEDISLRNGAV--VQNNG--SRPA			392
QY	416	GFVETSGHDLFKIDNAIVDAKEWLLDFDNVSINAEDPLF-----NNTGIND--EPTGTGGEA			470
DB	393	TTVLNAGHLNVLIESKTNVN-----NAKGATLSADGRTVIKEASITGTITVY			439
QY	471	SDPKKNSL-----KTLTNTTISN--YLKNAWTMTNITASRKL-----TVNSIN			513
DB	440	SSSKGNAELGNTRITGADVTVLSNGTISSSAVIDAKDTAHLAEGKPLSLEASTVTSDIR			499
QY	514	IGSNSHLILSHKGGGQVQIDGDTSK-----GGNLTIVSGGWDVHKNTLIDQGFNL			567
DB	500	LNNGS-----IKGGKQALLADDNITAKTNTLNPGLNYVHTG--KDLNLNVDKD-----			547
QY	568	ITAASVAFEGGNNKARDAANAKIYAQGVTVITGECKDFRANNVSLNGTGTGLNIISVNN			627
DB	548	LSASIIHLKSDN-----AAH-----ITGTSKLTASK--DMGVEAGSLNVTNT--N			589
QY	628	LTHNLSGTINI-----SGNITINQTRKNTSYQTSQSHDNVSNVSNALNLETGANFTFIKYISS			684
DB	590	LRTN-SGNLHIOAAKGNQL--RNTKLN-----AKALETTA-----LOG			626
QY	685	N--SKGLTQYRSSAGVNFNG-----VNGNMSFNLKEGAKVNEFKPKNNMNTSKPLPIR			737
DB	627	NIVSDGL-----HVSADGHVSLLANGNADET-----GHNTLTAKA-----			662
QY	738	FLANITATGGGVFDIYANHSGRGAELKMEISINSGANFTLNSHVRGDDAFKINKDLT			797
DB	663	-----DVNAGSVGKG--RLKADNTNITSSG-----DIT			689
QY	798	INATNSFSLRQTKDDFDYGVARNAINSTY--NISILGNVTLGQCN--SSSITGNITIEK			855
DB	690	LVAGNG-----IQLDGQKORSKNGKHISIKNNGNADLKNLNVHAKSGALNTHSDR			741
QY	856	A----ANVTLEA--NNAPNOQNIR-----DRVIKLGSLLVNG			886

DB	742	ALSIENTKLESTHNTLHNAQHERVTLNQVDAYAHRRHLSITGSIQWONDKLPSANKLVANG	801
QY	887	SLSLTGENADIKGNLTISASATFKGKTRTLNITGNFTNNGTAENITOG--VVLGNVT	944
DB	802	VIALNARYSOIADNTTLRAGA-----INLTAGTALVVRGNT--	837
QY	945	NGDGLNITTHAKRNQBSIIGGDIINKKGSINIT-----DSNNDAEIOIGGN--ISORE	995
DB	838	NMSTVTKTLEDNAELKPLAGRLNIEAGSGTLTIEPANRISATHTDLSIKTGKLLLSAKG	897
QY	996	GNLTSSDKINIKQITIKKIGDGEDS--SSDATSNANITIKTKELKLTEDLSISGFNFK-	1052
DB	898	GNAGAPSAQVSSLEAKGNIRLVTGETDLRGSKITAGKNLVWATTKGKLNIEAVNNSFSNY	957
QY	1053	-----AETAKDGRDLTIGNSNDGSGAEAKTVTFNNVKDSKISADGHNVTLNSKVTS	1106
DB	958	FPTQKAAELNQK--SKELEQOIAQLKKSPPKSLIP--TLQERDRFLAFYQAIKEVKGK	1014
QY	1107	SSNG-----GRESNNDTGLTITAKNVEVN-KDITSLKTVNITASEKVTITAGSTI---	1157
DB	1015	KPKGKEYLOAKLSAQNID--LISAGGIEISGSDITASKKLNLAAGVLPKAADSEAAAI	1071
QY	1158	-----NATNGKASITTKG-----DISGTISG-----	1179
DB	1072	LIDGITDQYEIGKPTTKSHYDKAALNKPRLTGRGTGVSHAAALDDAKIIIGASEIKAP	1131
QY	1180	NTVSVSATVDLTTRKSGS-----KIEAKSG-----EANTVSA	1210
DB	1132	SGSIDIKAHSDIVLEAGQDAYTFLTKTKGSKIIIRKFTSTTRDHLIMPAPVELTANGI	1191
QY	1211	TGTIGGTISGNTVNTANAGDLTVGNCAEIN--ATEG-----AATLTATGNTLTTEAGS-	1263
DB	1192	TLOAGNIEANTTRFNAPAGKVTLVAGELOLLAEEGIHKHELDVQKSRFIIKVGKSN	1251
QY	1264	-----ITSTKGQVDDLQAQNGSIAGSINAANVTNTTGTTLTVAGSDIKATSGTLVINAKD	1318
DB	1252	YSKNELNETKLPVRVVAQT--AATRSQWDTVLEGTPEKTTLAGADIQAGVGEKA--RAD	1306
QY	1319	AKL-----NGDASGDSSTEVA--NASSGSGSV--TAATSSSVNITGDLNVTNGLNI--	1365
DB	1307	AKIILKGIYNRIOSEKLETNSTVMQKQAGRGSTIETLKLPSFESPTPKLTAPGGYIVD	1366
QY	1366	ISKDGRNTVRLRGKEVEKVIOPGVASVEEVIEAKRV---LEKYKDLSDDEERETLAKIG	1421
DB	1367	IPKGNLKT-----ELEKLAKQPEYAYLKLQLOVAKNVNHNQVQLAYDKWDYKQEBGLTRAG	1420
QY	1422	VSAVREV	1428
DB	1421	AAIVTII	1427

Search completed: March 24, 2003, 15:24:34
Job time : 52 secs

